DOCUMENT RESUME

ED 057 675 48 FL 002 769

AUTHOR	Jackson, Kenneth L.; Whitman, Randal L.
TITLE	Evaluation of the Predictive Power of Contrastive Analyses of Japanese and English. Final Report.
INSTITUTION	Hawaii Univ., Honolulu.
SPONS AGENCY	Institute of International Studies (DHEW/OE), Washington, D.C.
PUB DATE	Sep 71
CONTRACT	OEC-0-70-5046-823
NOTE	126p.
EDRS PRICE	MF-$0.65 HC-$6.58
DESCRIPTORS	Applied Linguistics; Cloze Procedure; Comparative Analysis; *Contrastive Linguistics; *English; English (Second Language); Error Patterns; Interference (Language Learning); *Japanese; Language Research; *Predictive Ability (Testing); *Second Language Learning; Statistical Analysis; Syntax; Tables (Data); Test Construction; Test Reliability

ABSTRACT

This study tests three aspects of the problem of validity of contrastive analysis as a means for predicting errors or problems for second language learners: the constancy of foreign-language errors, the objectivity of the methods and procedures of contrastive analysis, and the capacity of contrastive analysis to make accurate predictions. Japanese students learning English provide the material for the investigation. The report contains details on the language testing and language sampling devices, the pattern of Japanese errors, and the four contrastive analyses used to predict difficulty. Statistical data on the success of the predictions are presented. The study concludes that the present state of contrastive analysis is inadequate, but with further research into the theory of interference contrastive analysis could play an important role in the theory of problem causation. A list of references is included. Appendixes present specific details on the construction of the experiment. (VM)

ED0 57675

Final Report

Contract No. OEC-0-70-5046 (823)

EVALUATION OF THE PREDICTIVE POWER
OF
CONTRASTIVE ANALYSES OF JAPANESE AND ENGLISH

Kenneth L. Jackson
and
Randal L. Whitman

University of Hawaii
Honolulu, Hawaii 96822

September 1971

The research reported herein was performed pursuant to a contract with the Office of Education, U.S. Department of Health, Education, and Welfare, under provisions of Title VI, Section 602, of the National Defense Education Act, Public Law 85-864, as amended.

FL002769

U. S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Office of Education
Institute of International Studies

U.S. DEPARTMENT OF HEALTH, EDUCATION
& WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

TABLE OF CONTENTS

Page

I. SUMMARY 1

II. INTRODUCTION

2.0 Introduction 3
2.1 The Dispute over Contrastive Studies 3
2.2 The Attack on the Linguistic Foundations of Contrastive Studies 5
2.3 The Attack on the Psychological Foundations of Contrastive Studies 6
2.4 Purpose of this Study 10
2.5 Limitations of this Study 11
2.6 Methods 11

III. THE LANGUAGE SAMPLING DEVICE AND THE LANGUAGE SAMPLING TEST

3.0 Introduction 13
3.1 The Testing of Syntax 13
3.2 Structures to be Measured 14
3.3 Principles Governing the Test Construction . . .16
3.4 Sampling 18
3.5 Administration of the Tests 24
3.6 Scoring the Tests 24
3.7 Validity and Reliability of the Tests 25
3.8 Intercorrelations of the Tests 32
3.9 Summary of the Chapter 34

IV. THE PATTERN OF JAPANESE ERRORS

4.0 Introduction 36
4.1 The Horizontal Variable 36
4.2 The Horizontal Pattern 43
4.3 The Vertical Variable 52
4.4 The Vertical Pattern 61
4.5 Summary of the Chapter 62

V. THE CONTRASTIVE ANALYSES

5.0 Introduction 67
5.1 Selection of the Contrastive Analyses 67
5.2 Variables in the Predictions of Difficulty . . .69
5.3 Sentence Items 73
5.4 Sentence Analyses 74
5.5 Predictions 75
5.6 Measures of Internal Consistency and Objectivity 76
5.7 Summary of the Chapter 78

Page

VI. THE POWER OF CONTRASTIVE ANALYSES TO PREDICT DIFFICULTY

6.0 Introduction .80
6.1 The Indices .80
6.2 Conclusions .84
6.3 Summary of the Chapter 84

VII. CONCLUSIONS AND RECCOMMENDATIONS 86

REFERENCES 89

APPENDICES 93

A. The Language Sampling Test 93
B. The Language Sampling Device 103
C. Test Construction 113
D. Optional/Obligatory Decisions on the LSD Sentences 120
E. The English Sentences and Their Japanese Translations 122
F. Linguistic Analysis of the Sentences for Contrast 125
G. Table of X-Values (Predictions of Level of Difficulty) 127
H. Table of e-Values (Gross Predictions of Difficulty) 129

LIST OF TABLES

Page

3.1 Distribution of Items on the LST and LSD According to Syntactic Structures 15

3.2 Means, Standard Deviations, Standard Errors of the Mean, and Correlation Coefficients of the Performance of the American Subjects with the Japanese H-Groups 26

3.3 Means, Standard Deviations, Split-Half Reliability Coefficients, and Standard Errors of Measurement for the LST and LSD by H-Groups 29

3.4 Average (Mean) Mean Scores, Standard Deviations, Split-Half Reliability Coefficients, and Standard Errors of Measurement of the LST and LSD for the H-Groups and V-Groups 30

3.5 Test-Retest Correlation Coefficients, Means, and Standard Deviations on HO2 Group Scores on the LSD 32

3.6 Intercorrelation Coefficients, Means, and Standard Deviations for the LST and LSD by Groups 33

3.7 Average (Mean) Intercorrelation Coefficients, Means, and Standard Deviations for the LST and LSD for H-Groups and V-Groups 34

4.1 Rank Order, Means, Standard Deviations, and Standard Errors of the Mean on the LST and LSD for the H-Groups 37

4.2 Intercorrelations of the LST Profiles of Performance for the H-Groups 38

4.3 Intercorrelations of the LSD Profiles of Performance for the H-Groups 41

4.4 Intercorrelations of the LST and LSD Profiles of Performance for the Three Lowest H-Groups: HO3, HO2, HO9 42

4.5 The Horizontal Pattern, P, of the Errors of Japanese Students in Rank Order from Least to Most Difficult for the LST and LSD 49

Page

4.6 Means, Standard Deviations, and Standard Errors of the Mean on the LST and LSD for the V-Groups 52

4.7 Rank-Difference Correlation Coefficients and Variance for the Rankings of Percentage of Difficulty on the LST and LSD for the American and V-Groups 53

4.8 Rank-Difference Correlation Coefficients and Variance for the Rankings of Percentage of Difficulty on the LST and LSD for the H-Groups and V-Groups 57

4.9 The Tenacity Factor, T, in Rank Order from Least to Most Tenacious for the LST and LSD . . . 63

5.1 Levels of Difficulty According to Stockwell . . . 72

5.2 Modified Stockwell Levels of Difficulty 73

5.3 Level of Difficulty Predictions (X) for each Contrastive Analysis, by the Investigators. . . . 76

5.4 Rank-Order Correlation Coefficients between the Investigators' Predictions of Difficulty Levels 77

6.1 Gross Capacity of Prediction 81

6.2 Relative Gross Capacity of Prediction 82

6.3 Capacity to Predict Relative Difficulty 82

6.4 Capacity to Predict Tenacity 83

6.5 Contrastive Analysis Performance Points 83

C.1 Initial Blueprint of the Tests 114

C.2 Initial Blueprint of the Tests Defined in Word-Class Categories 115

C.3 Number of Items on the LSD Accepted on the Basis of Native Speaker Performance 117

D.1 Expected and Actual Numbers of Marking Triplets . 120

Page

G.1 Predictions of Level of Difficulty 127

H.1 e-Values (Predictions of Gross Difficulty) 129

LIST OF FIGURES

Page

2.1 Osgood Transfer Surface 7

3.1 Location of Schools on a Map of Honshu 21

4.1 Performance Profiles of the H01 and H08 Groups on the LST 39

4.2 Horizontal Pattern of the H-Groups on the LST 45

4.3 Horizontal Pattern of the H-Groups on the LSD 47

4.4 Vertical Pattern of the V-Groups on the LST 55

4.5 Vertical Pattern of the V-Groups on the LSD 55

4.6 V-A and V-H Variance Characteristics 59

ACKNOWLEDGEMENTS

This project could never have been completed without the long hours put into it by Peter C. Lincoln, David Mordaunt, and Hildebert van Buren, all graduate students in the Department of Linguistics and Research Assistants in this project. As part of their duties, they assembled a twenty-seven page bibliography of works dealing with contrastive linguistics (barring phonological studies), which was given out to participants at the Pacific Conference on Contrastive Linguistics and Language Universals, University of Hawaii, January 1971.

All computational work was done at the University of Hawaii Computer Center by Wayne A. Fox, graduate student and Graduate Assistant in the Department of Psychology. Mr. Fox was especially invaluable in helping the Faculty Associates plan the nature of the analysis of the test performance, as well as in interpreting the results therefrom.

Thanks are also due to the Pacific and Asian Linguistics Institute, both in the person of the Director, Donald M. Topping, without whose help the project would never have reached formulation, and in the persons of the institute staff, who labored greatly with all clerical and secretarial work associated with this project.

Special mention should go to Paul Kakugawa, Fiscal Officer of the Projects and Grants Office, University of Hawaii, for clearing up our fiscal problems as we created them.

Finally, this project could not have been undertaken without the cooperation of the schools at which the tests described within were administered. We should like to thank especially the Principal, the Chairman of the English Department, and our "contact" within the English Department of each school (in this order, as listed below). (Our "contact" in each case participated in a special East-West Center program at the University of Hawaii in the year 1969-70).

- Waseda University Senior High School, Tokyo: Mr. Kôichi Okada, Mr. Hajime Haruki, and Messrs. Kohei Kojima and Yoshinori Endo.
- Waseda Business Junior High School, Tokyo: Mr. Hiroshi Takinuchi, Mr. Takesuke Takato, and Mr. Yoichi Tsuji.
- Minami Katsuhika Senior High School, Tokyo: Mr. Iwao Matsubara, Miss Junko Fuji, and Mr. Yoshiaki Yamada.
- Urawa Technical Senior High School, Urawa: Mr. Minoru Toida, Mr. Hajime Itakura, and Mr. Toshio Ehara.
- Kumagaya Senior High School, Kumagaya: Mr. Masayoshi Tsukagoshi, Mr. Kenichi Okada, and Mr. Yasuo Yokota.

- Fudôka Senior High School, Kazo: Mr. Rinji Yokotsuka, Mr. Morikazu Kawabe, and Mr. Hideo Nakazawa.
- Okayama Sozan Senior High School, Okayama: Mr. Kiyômi Egawa, Mr. Tadanori Yabuki, and Mr. Masayuki Miyake.
- Okayama Asahi Senior High School, Okayama: Mr. Wataru Goto, Mr. Jiro Fujiwara, and Mr. Shigeyoshi Shimo.
- Fujishima Senior High School, Fukui: Mr. Yô Ômori, Mr. Katsuyoshi Maegawa and Mr. Hisao Horiguchi.
- Nishi-Maizuru Senior High School, Nishi-Maizuru: Mr. Yaichiro Otsuki, Mr. Masao Takahashi, and Mr. Masao Doi.
- Ibaraki Senior High School, Mito: Mr. Seichi Shibata (Vice-Principal), Mr. Tsutomo Izumisawa, and Mr. Munetsugu Uruno.
- Toyama Senior High School, Tokyo: Mr. Fumio Ikeda, Mr. Kaneyuki Kobayashi, and Mr. Seichi Saida.

We should also like to thank the many Japanese teachers who helped administer the tests in their classrooms.

We should also like to thank Mr. Charles W. Popovich, Assistant Superintendant of Schools, Sheridan Public Schools, Sheridan, Wyoming for his help in arranging to have the tests administered. Also, the Principals and teachers of Coffeen, Highland Park, Taylor, and Linden Elementary Schools, Sheridan, Wyoming.

Honolulu, Hawaii
September 1971

Kenneth L. Jackson
Randal L. Whitman

CHAPTER I. SUMMARY

The purpose of this project, pursuant to contract number OEC-0-70-5046 (823) with the Office of Education, Department of Health, Education, and Welfare, was to examine and evaluate the capacities of certain contrastive analyses of Japanese and English to predict errors in English on the part of Japanese who are learning English. It fell also within the scope of this project to determine whether or not the nature of second-language errors is such that an 'ideal' contrastive analysis could fruitfully be expected to predict them.

The following procedures were involved:

1. Construction of two tests of English: the Language Sampling Device (LSD), a Cloze-procedure test, and the Language Sampling Test (LST), a multiple-choice test.

2. Administration of both tests to approximately 4000 Japanese students in sixteen groups - ten groups of tenth graders spread throughout Japan, and the seventh through twelfth grades of one school. Also, to approximately 200 fourth grade students in several elementary schools in Wyoming.

3. Computer analysis of the test results to determine the relationships the different Japanese groups and the relationships between the Japanese and Wyoming groups.

4. Extraction of predictions of difficulty on the tests from four contrastive analyses.

5. Computer analysis of the relationship between the performance data and the predictions.

The significant findings may be summarized as follows:

Of the four analyses, none was capable of predicting Japanese performance on the tests, at least in terms of the analyses themselves. Only one analysis achieved results at a significant level (but sporadic and minor), but then only if the predictive system were "reversed" - i.e., if the predictive hierarchy of difficulty were reversed.

Analysis of the data reveals that there is a very high level of commonality in the problems that the Japanese students had on the tests, and that, furthermore, there is a strong suggestion that the native language of the students may play a considerable role in this commonality. This,

coupled with the negative findings given above, indicates that the present state of contrastive analysis is inadequate, but with further research into the theory of interference, contrastive analysis could play an important role in the theory of problem causation.

CHAPTER II. INTRODUCTION

2.0 "Contrastive analysis" is basically a cover term for any study which examines phenomena from two languages in terms of a common paradigm. As such, contrastive analysis is a necessary tool for the advancement of linguistic theory, since, in essence, linguistic theories are either developed from the point of view of adequacy in explaining facts in several languages, or, if developed in terms of one language, must be texted for adequacy on other languages. In either case, contrastive analysis is employed in a general sense of the term. In a more specific sense, however, contrastive analysis is understood to be a form of analysis by which language-learning problems are isolated, as a function of the differences in the way the languages work. In this project, contrastive analysis in this latter, more specific meaning, will be evaluated.

2.1 The Dispute over Contrastive Studies.

The background of this study is rich in evidence that the theoretical bases of contrastive analysis are themselves in dispute.[1]

With the development of the more or less independent field of applied linguistics, and the 'linguistic' methods of teaching foreign languages, contrastive analyses of two languages began to play an important role in the development of materials for second language teaching. Ferguson (in Moulton 1962, v) for example, expresses the attitude toward the usefulness of contrastive studies that is held by a number of linguists.

> "The Center for Applied Linguistics, in undertaking this series of studies [The Contrastive Structure Series] has acted on the conviction held by many linguists...that one of the major problems in the learning of a second language is the interference caused by the structural differences between the native language of the learner and the second language. A natural consequence of this conviction is the belief that a careful contrastive analysis of the two languages offers an excellent basis for the preparation of instructional materials..."

---

[1] All comments are directed to Syntactic contrasts, although most discussion is also applicable to phonological contrasts.

To a certain extent, it even appears that such studies are the most important contribution that linguists make to foreign language teaching. Wilga Rivers (1964, 14) states,

> "Careful scientific analysis of contrasts...is of course the distinctive contribution of the linguistic scientists."

The works of Robert Lado (1957, 1961) set the pattern for many of the ensuing contrastive analyses. This early pattern might be characterized as involving:

(a) a strong emphasis on surface structure differences,

(b) the prediction of 'difficulty', sometimes specific 'errors', on the part of the language learner,

(c) the optional erection of a hierarchy of difficulty.

This type of output became the basis for the selection and organization of the content of a number of language teaching materials, such as those by Lado and Fries (1958) and by Paul Roberts (1963-1965). Many contrastive analyses themselves give explicit directions as to how to implement the analytic findings in the classroom (such as Stockwell, 1965).

More recently, as a result of the development of the transformational-generative school of linguistic theory, contrastive analyses have continued to flourish (especially in the form of Ph.D. dissertations) but with an increasing focus on the transformational structures of the languages rather than the surface structures. Dingwall (1964) may be considered one of the earliest theoreticians of transformationally based contrastive linguistics, and certainly one of the most rigidly transformational. His focus is unremittingly on the underlying structures of the two languages. Stockwell (1965), on the other hand, represents a way-station between the purely structural and the purely transformational approaches. While his basic philosophy is that of the generative grammarian, Stockwell's operational criteria in establishing a hierarchy of difficulty include surface structure phenomena.

In general, the advent of generative linguistics did not itself lead to a loss in confidence in contrastive studies. Of course, it did lead to argument as to the greater effectiveness of contrasting languages described one way versus the other.

Some (but certainly not all) of the contrastive analyses were tested in one way or another. Kleinjans (1959) and Jackson (1970), for example, included extensive tests which supported their respective predicted hierarchies. Stockwell (1965) and Hashimoto (1966), on the other hand, did not.[1] It can be claimed, in any case, that any supporting tests were designed to find those problems that were predicted, and so might be biased in favor of the analysis. Then, too, many of the tested analyses were quite limited in scope - Kleinjans to noun-head modification patterns, Jackson to middle adverb patterns - that even if their tests were adequate, it is difficult to generalize the value of their approaches to other areas of the language. The empirical validity of contrastive analysis has not been conclusively demonstrated in any case.

2.2 The Attack on the Linguistic Foundations of Contrastive Studies.

Parallel to the continuing development of contrasive analyses, there has been a swelling theoretical attack on the linguistic foundations of contrastive studies.

One of these attacks (cf. Ritchie, 1967) is that the linguistic structures of one language cannot be sensibly compared to those of another unless the two are closely related. In principle, this argument follows from an assumption that it makes little sense to talk about transfer and interference when the structures concerned are grossly disparate (as between unrelated languages), whereas when the structural systems are grossly similar, then points of difference may become significant.[2]

Another attack is that the methodology of contrastive analysis represents a vast --and quite uncharted-- leap from the hypothetical psychological problems of a second language learner to the linguistic mechanisms involved in mapping one linguistic structure onto another (cf. Ritchie, 1967). Wardhaugh (1970) goes so far as to state that contrastive theory makes "impossible demands" on linguistic theory.

Finally, the most common negative view is that contrastive analysis should not be predictive but explanatory, if it is to have any function at all (cf. Carroll, 1968, et al.). As Gradman (1971) says,

---

[1] Stockwell (1968) claims that his hierarchy was derived empirically and not theoretically, so perhaps testing is obviated in favor of his experience.

[2] This parallels the argument that Jakobovits (1970) raises concerning the use of the native language in teaching a related or unrelated target language.

> "It is only proper to argue for the relegation of contrastive analysis to a broader theory of error analysis, and indeed only a subcomponent of that theory."

2.3 The Attack on the Psychological Foundations of Contrastive Studies.

The psychological foundations of contrastive studies are no less questioned.

The concept of interference has its psychological home in the work first reported by Osgood (1949), as represented specifically in the "transfer surface" (Fig. 2.1).

The transfer surface was interpreted to illustrate two principles governing the transfer of learning from one learning operation to the next.

(1) As the similarity of the two stimuli decreases, transfer (of any kind) approaches zero - i.e., there is less and less effect of the first learning on the second.

(2) As the similarity of the two responses decreases, the amount of transfer decreases (ultimately to negative transfer, or interference).

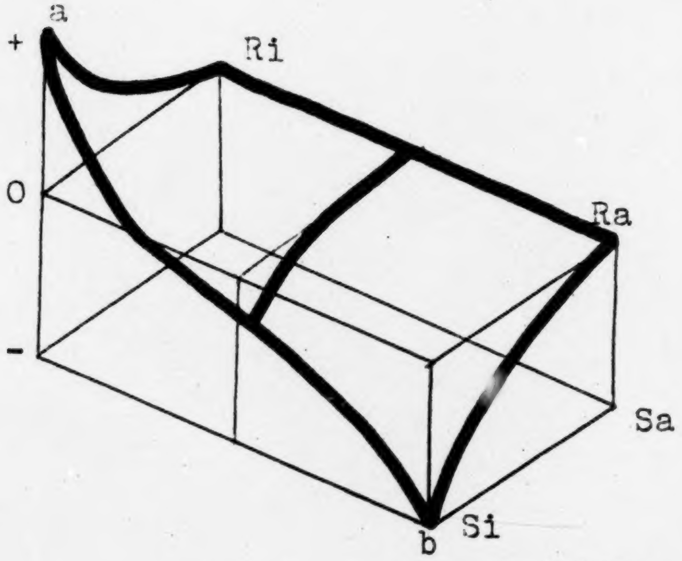
Kleinjans (1959) actually set out to test the validity of the transfer principles in his contrastive study of Japanese and English. In fact, one of his assumptions in this task was the validity of the contrastive analysis. In his study, he encountered the most significant problem in relating the model of transfer to linguistic data: defining stimuli and responses in terms of natural language.

In the way that the transfer model was derived, two tasks were related. Task 1 involved learning a particular response 1 to a particular stimulus 1, and Task 2 involved learning response 2 to stimulus 2. In the classical experiment, one set of subjects was given both tasks in sequence, and another set of subjects was given Task 2 alone. The two groups' rates of learning Task 2 were compared, and any difference between the two groups was attributed to the effect of the former group's having previously learned Task 1.

In a second language learning situation, Task 1 is presumed to be the learning of the students's native language, and Task 2 the learning of the target language. (No real control group is available, unless one conceives of the

Fig. 2.1

Osgood Transfer Surface



(from Kleinjans, p. 214.)

Vertical dimension: amount of transfer, with neutral transfer represented by the plane (O).
Left-to-right: amount of shift in RESPONSE similarity, from Response identity (Ri) to Response antagonism (Ra).
Front-to-back: amount of shift in STIMULUS similarity, from Stimulus identity (Si) to Stimulus antagonism (Sa).

Point (a) is the point of maximum positive transfer (facilitation), and point (b) is the point of maximum negative transfer, (interference).

control group as being the native speakers of the target language, and in this case, it must be further assumed that first and second language learning involve the same type of learning, an assumption that is hotly disputed in modern psychology).

In any case, the specificity of stimulus and response is lost to a great degree in language. Where Osgood (and others working with the transfer model) worked with specifiable stimuli and responses, Kleinjans (and others attempting to impose this model on language learning) was forced to adopt rather vague definitions for both stimulus and response. In his model of production, for example, a stimulus was seen to be a meaning that the speaker wished to express, and the response was the expression of that meaning. Kleinjans then posited that he could usefully study cases for which the desired meanings (stimuli) in the two languages were the same, such that difficulty would be a function of the difference in the expressions (responses) in the two languages (cf. the second transfer principle above).[1]

The Kleinjans and Jackson studies are not the only two syntactic studies which attempt to make direct use of this model of transfer. Politzer (1960) refers to "interference" as a general concept without relating it to any specific model, but in a later work (Politzer, 1965), he refers directly to the model discussed above. Other studies make no reference to interference at all, presumably making it axiomatic that structural differences lead to language learning difficulty.

Whatever its merits with respect to language learning, the Osgood model of transfer is rooted in the general behaviorist approach to psychology. Parallelling the theoretical breach between the structural and the generative linguists, there is a growing school of psychologists who have rejected the behaviorist model in favor of a cognitive model as applied at least to the processes of language learning. Cognitive psychologists, as they reject the behavioral model, reject also the transfer model as not being appropriate to the language learning situation. No cognitive model of control group as being the native speakers of the target language, and in this case, it must be further assumed that first and second language learning involve the same type of learning, an assumption that is hotly disputed in modern psychology).

---

[1] It is interesting to note that Kleinjans perceived the desired meaning (the stimulus) as being independent of the variable MEANING, one of the descriptors of an expression. Jackson (1970), on the other hand, saw the variable MEANING and the stimulus as being identical, with a resulting hierarchy that differed markedly from Kleinjans'.

In any case, the specificity of stimulus and response is lost to a great degree in language. Where Osgood (and others working with the transfer model) worked with specifiable stimuli and responses, Kleinjans (and others attempting to impose this model on language learning) was forced to adopt rather vague definitions for both stimulus and response. In his model of production, for example, a stimulus was seen to be a meaning that the speaker wished to express, and the response was the expression of that meaning. Kleinjans then posited that he could usefully study cases for which the desired meanings (stimuli) in the two languages were the same, such that difficulty would be a function of the difference in the expressions (responses) in the two languages (cf. the second transfer principle above).[1]

The Kleinjans and Jackson studies are not the only two syntactic studies which attempt to make direct use of this model of transfer. Politzer (1960) refers to "interference" as a general concept without relating it to any specific model, but in a later work (Politzer, 1965), he refers directly to the model discussed above. Other studies make no reference to interference at all, presumably making it axiomatic that structural differences lead to language learning difficulty.

Whatever its merits with respect to language learning, the Osgood model of transfer is rooted in the general behaviorist approach to psychology. Parallelling the theoretical breach between the structural and the generative linguists, there is a growing school of psychologists who have rejected the behaviorist model in favor of a cognitive model as applied at least to the processes of language learning. Cognitive psychologists, as they reject the behavioral model, reject also the transfer model as not being appropriate to the language learning situation. No cognitive model of interference has been suggested to replace the transfer model, with the result that contrastive analysis has been bereft of its psychological basis (such as it was) in the eyes of most contemporary psycholinguists.

---

[1] It is interesting to note that Kleinjans perceived the desired meaning (the stimulus) as being independent of the variable MEANING, one of the descriptors of an expression. Jackson (1970), on the other hand, saw the variable MEANING and the stimulus as being identical, with a resulting hierarchy that differed markedly from Kleinjans'.

With psychologists denying the psychological validity of interference, and with linguists denying the linguistic validity of the structural-difference/difficulty axiom, one can see why contrastive analysis should be in such dubious straits.

2.4 Purpose of this Study.

It would appear, then, that there is a need for an objective study of the validity of contrastive analysis as a means for predicting errors or problems for second language learners. The investigators proposed to test three aspects of this problem of validity: the constancy of foreign-language errors, the objectivity of the methods and procedures of contrastive analysis, and the capacity of contrastive analyses to make accurate predictions.

To test the constancy of foreign-language errors, the investigators collected a sample of the English syntactic usage of groups of Japanese students. These groups consisted of Japanese who were learning English under varying conditions of time and space, i.e., different levels of study, teachers, textbooks, geographical areas, socio-economic backgrounds, etc. The investigators then determined whether or not there was any significant difference in the patterns of responses between and within these groups.

To test the objectivity of the methods and procedures of contrastive analysis, the investigators attempted to extract from certain contrastive analyses predictions of difficulty for specified test items to determine to what degree the procedures of the analyses are objective and consistent.

To test the capacity of contrastive analyses to make accurate predictions, the investigators gathered the predictions from selected contrastive analyses based on differing models of language description and/or differing methods of contrastive analysis, and compared these predictions with the errors found in the sample of Japanese English performance.

The output of this study is a set of indices which are used to evaluate a given contrastive analysis in terms of its predictive capacities. The details of the indices are dealt with in Chapter 6.

2.5 Limitations of this Study.

Since it would have been beyond the time limits imposed on this study as well as the capabilities of the investigators to attempt to test the validity of all extant contrastive analyses between all languages, the study was limited to four contrastive analyses of English and Japanese.

The Japanese and English languages were selected for this study for three reasons. First, extant contrastive analyses of these two languages based on differing models of language description and/or models of contrastive analysis were readily available in sufficient numbers to make the study practicable. A partial list of contrastive studies of Japanese and English, for example, numbers twenty-six, of which fifteen are phonological, ten syntactic, and one semantic in nature. Secondly, both of the investigators had themselves completed contrastive analyses of certain aspects of these two languages. Thirdly, the services of thirteen Japanese senior high-school teachers of English who studied at the University of Hawaii in the year 1969-1970 were available for aid in arranging for the administration of the language sampling tests in Japan.

2.6 Methods.

The procedures followed in this investigation consisted of the following steps:

Step I. Collecting a sample of the English usage of 180 fourth-grade American informants; a set of eleven groups each composed of 200-250 tenth-grade Japanese informants; a set of six groups each composed of 180-250 Japanese informants ranging from the seventh to the twelfth grade of one school complex.

Step II. Determining the patterns of error between and within the groups of each set of Japanese informants and the patterns of difference between the American and Japanese informants.

Step III. Gathering the predictions of a set of four contrastive analyses of English and Japanese which represent different descriptive bases and/or different contrastive methodologies.

Step IV. Measuring the degree of subjectivity of each of the contrastive analyses gathered.

Step V. Comparing the predictions of the contrastive analyses gathered with the English performance of the Japanese informants.

Step VI. Ranking each of the contrastive analyses with respect to a series of indices representing different aspects of predictive capability.

The details of the results obtained by Step I are reported in Chapter 3, for Step II in Chapter 4, for Steps III and IV in Chapter 5, and for Steps V and VI in Chapter 6.

CHAPTER III. THE LANGUAGE SAMPLING DEVICE AND LANGUAGE SAMPLING TEST

3.0 The Language Sampling Device and Language Sampling Test were developed to provide a standardized procedure for collecting a sample of the English usage of Japanese and American students. They were designed to measure specific English syntactic structures for which no previously prepared tests were available.

This chapter is intended to give a general overview of the measurement philosophy underlying their development and some of the actual background which determined their content and structure.

3.1 The Testing of Syntax.

Since Lado's publication on language testing in 1961 (Lado, 1961), considerable progress has been made in the theory and practice of discrete point testing of syntax. Objective test items of the multiple-choice type in which the student is asked to choose from four choices the correct answer to fill in a blank have been developed. Such items generally yeild highly satisfactory item-test correlations and are thus highly reliable. They are regarded by many to be valid measures of syntactic knowledge of the foreign language, i.e., the ability to choose correct grammatical forms. Finally, they can be scored quickly. Although such items do constitute a valid sample of language behavior, they tend to draw on highly specific knowledge, rather than broad competence in the language as a whole. In other words, multiple-choice items of the type described do not allow the student to show how knowledgeable or ignorant he is of the syntactic structures and cues in the language. Carroll (1959.2) argues that such items "tend not to measure accurately at upper levels of ability, because they measure the mere existence of language habits rather than their strength."

In order to discover testing methods which might circumvent the current difficulties inherent in multiple-choice test procedures, Carroll and others (1959) made an extensive study of the use of 'cloze' (from 'closure') items for measuring achievement in foreign languages. Such items ask the subject to restore a mutilated text by supplying a word, letter, or phrase. In their investigation, the researchers found that 'cloze' items seem to be suitable for assessing group differences in second language competence, but not individual differences. Such items were found to be relatively unreliable and affected by various extraneous factors. While the items are simple to prepare, they are suitable only for testing written language and are cumbersome to score. The items appear to measure the ability to restore texts, an ability dependent to a considerable extent upon cognitive ability, reasoning ability, and ideational fluency. Although Carroll indicates these abilities are independent of "language competence", more recent studies of language acquisition and processing by psycholinguists suggest that cognitive ability is a vital element in the process of using and acquiring language competence (Chomsky, 1968a, MacIntyre, 1970, Bever, 1970). Unlike multiple-choice items, 'cloze' items do test the examinee's ability to produce utterances and place great demands on him to draw on his competence to select the appropriate word to fill the blank in the mutilated text. Also, Carroll (1959) reports that there is a significant difference between average performance between learners of a language and those who have achieved native language mastery, thus providing a measure of strength.

Since some of the weaknesses of the multiple-choice test items are compensated for by 'cloze' procedures, and vice versa, the most valid sampling of the syntactic usage of language learners would seem to be in the form of items using both methods of testing. The Language Sampling Device (LSD) and the Language Sampling Test (LST) were developed as equivalent forms of the same test of syntactic structures but differing in their methods of testing. The LST follows a multiple-choice format while the LSD follows a 'cloze' procedure.

3.2 Structures To Be Measured.

At the outset, it was necessary to decide which structures should be measured by the tests. It was obvious that a practical test of syntax could not conveniently include measures of all the known structures in the language. Therefore, the tests were limited to Noun-phrase structures, Verb-phrase structures, Negative structures, and Interrogative structures. Since the structural possibilities are far richer in Noun-phrase and Verb-phrase structures, the tests were weighted in their favor, having fifteen and fourteen items out of forty, respectively.

Even within these four areas of interest, the variety of possible structures is so great further limitations were imposed on what could be tested economically yet at the same time yield a profile of the English usage of Japanese and American students. Clues as to which structures could be tested was obtained through administration of experimental forms of the tests.

The tests finally developed include forty items. The distribution of the items according to syntactic structure is shown in Table 3.1. (Copies of the LST and LSD and the details of their construction will be found in Appendices

3.3 Principles Governing the Test Construction.

Major Considerations. In constructing the LSD and LST, certain practical criteria were established to govern the work. Some of these are presented in special sections below. The sampling, administration and scoring of the tests will be found in Sections 3.4, 3.5, and 3.6, respectively. Validity and reliability coefficients for the two tests will be found in Section 3.7. Section 3.8 presents the intercorrelations of the tests. In the following paragraphs, certain other governing principles are noted.

The Tests Should Be Universal. Harris (1969,12) defines a 'universal' test as "one that can be used with students of disparate language backgrounds." Since the sample obtained by the tests would be used ultimately to validate particular contrastive analyses, a universal test of syntax was mandatory in order to be free of any expectations of what Japanese performance would be like and thus skew the results in favor or against one or another of the selected contrastive analyses.

'Universality' of test design was felt to be a matter of ensuring minimum bias toward pre-selected areas of expected difficulty. Wherever possible, the selection of test items followed strongly randomized procedures, involving, for example, a random numbers table. Sentences containing the selected tokens were obtained from the entry for the token in Thorndike Century Junior Dictionary, West's General Service List, or The Random House Dictionary of the English Language which were each consulted in turn until sentences were obtained for each of the selected tokens.

The three distractors for the LST were selected on the basis of the word-class category being tested. If, for example, the category was one of verbal tense, the distractors were various tenses. If, however, this process did not yeild a sufficient number of distractors, others were selected either from errors made by native speakers of English on the experimental edition of the LSD or on an arbitrary basis. In sum, no attempt was made to select items or distractors on the basis of the fact that Japanese students were going to take the tests. While this does not guarantee the 'universality' of the test (a matter that can only be judged by sampling and analysis), at least it is not a test specifically for Japanese.

The Tests Should Measure Syntax. Given present methods of testing, a 'pure' test of syntax is virtually impossible because syntax cannot be manipulated without vocabulary. Thus all tests of syntax are to a greater or lesser degree a test of vocabulary, depending on how well-known the

A, B, and C, respectively).

TABLE 3.1

Distributions of Items on the LST and LSD According to Syntactic Structure

Structures	Number of Items	Sub-Total	Total Items
I. NOUN-PHRASE STRUCTURES			15
A. Noun-head Structures		3	
1. Count Noun, singular	1		
2. Count Noun, plural	1		
3. Mass Noun	1		
B. Left-branching Modification Structures		6	
1. Adjectives	3		
2. Determiners	3		
a. Determiners	2		
b. Pre-determiners	1		
C. Right-branching Modification Structures		3	
1. Relatives	3		
D. Noun-phrase Substitute Structures		3	
1. Pronouns	2		
2. Function Nouns	1		
II. VERB-PHRASE STRUCTURES			14
A. Main Verb Structures		6	
1. V-Ø	1		
2. V-ed	3		
3. V-en	1		
4. V-ing	1		
B. Auxiliary Structures		6	
1. +V-Ø	1		
2. +V-ing	1		
3. +V-en	3		
4. + to V	1		
C. Verb Complement Structures		2	
1. + to V	1		
2. +V-ing	1		
III. NEGATIVE STRUCTURES			4
IV. INTERROGATIVE STRUCTURES			7
A. Wh-Questions		3	
B. Regular Questions		4	

vocabulary is for the examinee. To reduce any interference in the testing procedure from vocabulary unknown to the informant, only words which ranked in the first 500 words of highest frequency on the Kucera-Francis list (1967) were included in the sentences in the tests. Since frequency on the Kucera-Francis list is determined solely on the basis of graphological form, further interference might arise from the lexical meaning of a given word. To compensate for this factor, a semantic index for each of the tokens was obtained by multiplying the frequency of occurrence on the Kucera-Francis list by the semantic frequency given in West's General Service List. If the resulting product was greater than 194, the token was defined in that meaning.[1] Thus, through controls on the frequency of occurrence and semantic frequency, items on the test were obtained which were hoped to be relatively free from interference by unknown vocabulary or lexical meanings. Since some testees were just beginning their English studies, of course, this hope was entertained primarily for the third-year-of-English students, who comprise the bulk of the testees.

The Tests Should Measure Power. The purpose of collecting the sample of English usage was to discover the level at which the examinee could perform, rather than how rapidly he could respond. Since 50 minutes were allowed for administering each of the tests, experimental versions of the LSD with 50 items were tried out on both 3 English-speaking children (aged 8, 9, 9 1/2) and 3 Japanese 7th graders, to determine how long the test would take. The longest time registered was 26 minutes for a native speaker of English and 35 minutes for a Japanese student. It was felt, then, that most of the students tested would have little difficulty in completing the 40 item tests in the time allowed. In fact, most of the Japanese subjects completed either test in about 30 minutes.

The Test Materials Should Be Practical. The two tests were printed in separate test booklets with directions in Japanese for the Japanese students and in English for the American students so the tests could be administered in separate periods.

Students recorded their responses for the LST on IBM #503 answer sheets, which were machine scored. They recorded their responses to the LSD in the test booklet. These responses were later hand scored and converted to IBM format.

---

[1] The number 194 was selected arbitrarily. It represents the lower limits of the first 500 words on the Kucera-Francis list.

The Tests Should Be Easy to Administer. The basic directions for taking each test were printed in the test booklet. These were read aloud by the administrator and silently by the persons being tested. Section 3.4 of this chapter contains information concerning the general testing conditions and gives the details regarding the administration of the tests. The two governing principles in planning the administrative features of the test were that they should be administratable by the classroom teacher with a minimum of special training and the directions and examples should be clear to the examinees. Experimental verions of the directions were tried, and an actual tryout by a tester with a minimum of 'briefing' was conducted.

The Two Tests Should Be Equivalent Forms. Since the only difference desired between the LST and LSD was the method of testing, steps were taken to ensure that they were equivalent in content and significance. Forty acceptable items were obtained from experimental versions of the LSD, and placed in random order. This final edition of the LSD was used as the format for the LST. The same word-class categories and order of items was observed. If for any category, there was a usable item which did not appear on the final form of the LSD, it was used as the basis for the equivalent item on the LST. This process yielded twelve of the LST items. For the remaining eighteen items, the sentence stem from the LSD was used and words were substituted into the sentence frame in keeping with the co-occurrence restrictions of the key word which was preserved. In this way, all forty of the LST items were obtained which matched those of the LSD in content and significance.

3.4 Sampling.

Using the LST and LSD, samples of English usage were obtained from the following groups:

Group A:	Fourth-grade American students.
Groups H01....., H11	Eleven groups of tenth-grade Japanese students.
Groups V07....., V12	Six groups of Japanese students in successive years of English study, beginning with the seventh grade and ending with the twelfth grade.

(NOTE: It should be noted that one H-group, H01, is the same as the tenth grade V-group, V10).

Each of these groups is described in turn below.

Group A was drawn from the Sheridan Public Schools, Sheridan, Wyoming. It was chosen because it is a fairly representative American school system, and because the area is not noted for having any particular regional dialect traits. The fourth grade was chosen because students at this level may be assumed to be competent native speakers of English, capable of reading and writing (i.e., of taking the LSD and LST). Yet, they can also be assumed to have had only a minimal amount of formal instruction in English grammar. The fourth graders were boys and girls eight and nine years of age in six different classes. 173 students took the LST and 181 took the LSD in January, 1971. The purpose of this particular sample was to obtain a measure of the intrinsic difficulty of the measuring instruments themselves.

Groups H01,..., H11 came from eleven different senior high-schools located in Tokyo, Kyoto, Saitama, Okayama and Fukui prefectures. The tenth-grade was chosen because it is the lowest year of instruction in the Japanese senior high school system and the third year of English instruction. The schools are located on the map of Honshu, Figure 3.1. The numbers refer to:

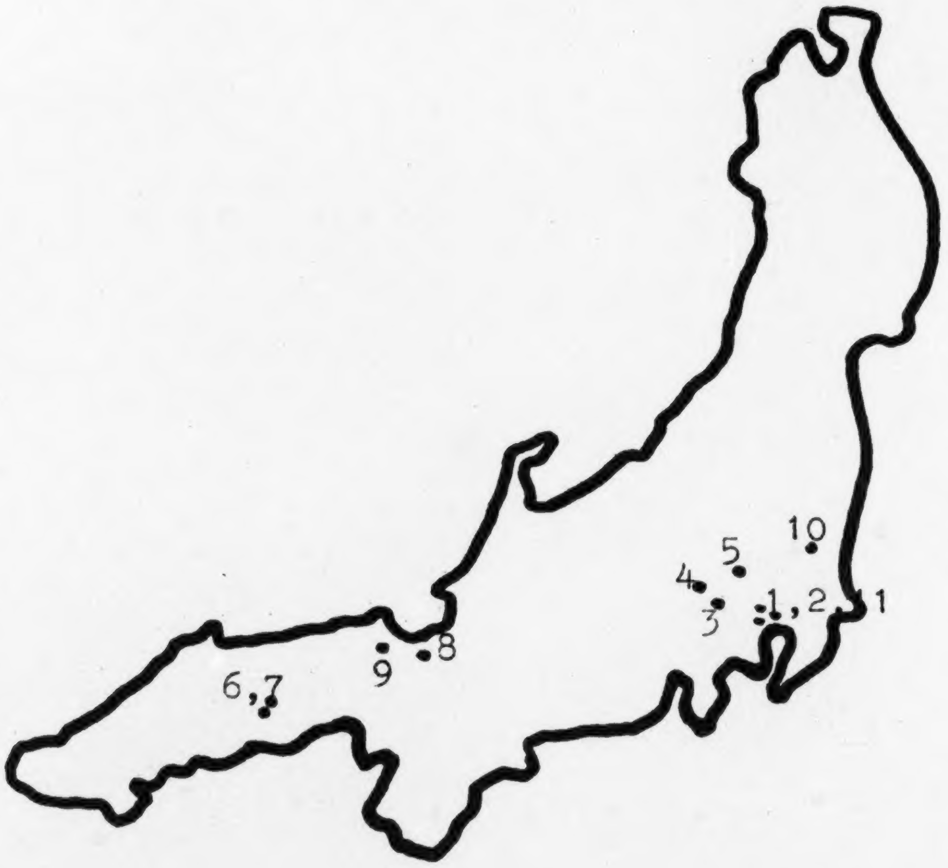
H01	Boys private school, Tokyo.
H02	Co-educational prefectural school, Tokyo.
H03	Co-educational prefectural technical school, Tokyo.
H04	Co-educational prefectural school, Kumagaya.
H05	Co-educational prefectural school, Koza.
H06	Co-educational prefectural school, Okayama.
H07	Co-educational prefectural school, Okayama.
H08	Co-educational prefectural school, Fujishima.
H09	Co-educational prefectural school, Maizuru.
H10	Co-educational private school, Mito.
H11	Co-educational prefectural school, Tokyo.

The schools represent four dialect areas: Eastern Kanto (H10), Western Kanto (H01, H02, H03, H04, H05, H11), Kinki (H08, H09), and Chugoku (H06, H07) (Tokieda, 1955). All of the schools are located in or near major urban centers except three (H08, H09, H10) which are rural.

Although the syllabus of the English curriculum in Japan is determined by the Ministry of Education and is, therefore, roughly the same thoroughout the country, the teachers, their methods of teaching, and the textbooks they use vary from school to school, along with the socio-economic backgrounds of the individual students. The only uniformities among the eleven groups were that most of the students who began their study of English in the seventh grade, were fifteen and sixteen years of age, were native speakers of Japanese, and had learned English from textbooks based on the same general syllabus. All of the differences between these groups--geographic, dialectal, pedagogical, socio-economic, etc.--were designated as the HORIZONTAL

Fig. 3.1

Locations of Schools
on a map of Honshu



(Note: The numerical designations of the schools on this map have had the 'H' prefix dropped. Thus, school H01 is listed on the map simply as 1.)

VARIABLE (H) (whence comes the "H-groups.")

The purpose of this sample was to determine the degree to which this horizontal variable has a significant influence on the English performance of Japanese informants, if at all.

The total number of students in the H-groups who took the LST was 1896 while 2544 took the LSD. One group, H02, was retested with the LSD four weeks after the original testing, to determine test-retest reliability. Because of a lack of time, the LST was not administered to two groups, H06 and H09. When the tests were administered in October, 1970, the students had completed three and one-half years of English study.

Groups V07,..., V12 were drawn from the private feeder schools of a private university complex in Tokyo. V07, V08, and V09 designate the seventh, eighth, and ninth grades of the commercial junior high school, while V10, V11, and V12 designate the tenth, eleventh and twelfth grades of the boys senior high school.

In order to minimize the effects of the horizontal variable discussed above, these private schools were chosen because they are a single school complex in which the teachers, their methods of teaching, and the textbooks they use remain the same for considerable periods of time. Furthermore, this school complex is one of Japan's outstanding private schools and, therefore, draws a large number of high-calibre students from throughout the country, although most of the students are from Tokyo and the surrounding prefectures. For these reasons, then, a certain degree of uniformity was expected in the backgrounds of the informants of this group.

The purpose of this particular sample was to measure the degree to which particular errors are retained in time. The differences due to time were designated as the VERTICAL VARIABLE (V) (whence comes the "V-groups.")

The total number of students in the V-groups who took the LST was 1177 and for the LSD, 1221. When the tests were administered in October, 1970, the seventh graders had completed one-half year of English study; the eighth grade, one and one-half years; the ninth grade, two and one-half years; the tenth grade, three and one-half years; the eleventh grade, four and one-half years; and the twelfth grade, five and one-half years. The students ranged in age from twelve in the seventh grade to seventeen in the twelfth grade.

3.5 Administration of the Testing.

The tests were administered during two normal fifty-minute English class periods on separate days by the classroom teacher. The tests were administered by the classroom teacher because it was felt that the presence of an outsider, particularly in Japan, might obscured the performance on the tests. In the first period, the classroom teacher administered the LSD, reading the instructions and answering questions concerning the instructions, but offering no assistance with any of the items. In the second period, on the following day, the teacher administered the LST in a similar fashion.

The LSD was administered first in hopes that the range of answers on the 'cloze' items might be greater. While it was expected that there would be some practice effect on the performance on the LST given the following day, it was felt that this would be minimized because the tests were sufficiently long, i.e., forty items, and the stems of the items while testing the same thing in the same order, were different on the two tests. As soon as the tests were completed, they were returned to the investigators for scoring.

3.6 Scoring the Tests.

The LST was machine-scored while the LSD, because of its method of testing, was hand-scored. This hand-scoring was done by three research assistants who were to determine whether each of the responses to the forty items was grammatical or ungrammatical. Grammatical, ungrammatical and omitted items were indicated on an IBM answer sheet and tabulated by computer. All computation was done by the Computer Center, University of Hawaii.

Any item on the LST or LSD which had two answers was omitted from tabulation. No answer was given partial credit. Each item was considered to be answered correctly or incorrectly. An item on the LST was correct only if the correct answer had been filled in. An item on the LSD was correct only if the researcher judged it to be grammatical. No correction for guessing was made.

After the scoring of the LSD was completed, the objectivity of the scoring procedure was tested. The two investigators each scored separately a random sample of 480 items from the six V-groups which had been previously scored by a research assistant. The Pearson product moment correlation coefficient for each pair of scorers was as follows:

1. Jackson and Whitman	1.000
2. Jackson and Research Assistant	.998
3. Whitman and Research Assistant	.998

The coefficients indicate that the scoring methods were sufficiently objective.

3.7 Validity and Reliability of the Tests.

Earlier in this chapter it was noted that the LST and LSD were developed for the purposes of collecting a sample of the English usage of Japanese students. Tests were desired which would be appropriate as to the abilities measured, practical as to the format and administration and technically satisfactory as to reliability. But also tests were desired which were valid in content.

Content validity involves the inspection of the items in a test to determine if they contain language problems that are valid. It also involves the detection of extraneous factors which may invalidate the items. To control the interference of extraneous factors, a 100 item experimental form of the LSD was administered to 65 University of Hawaii freshman in July, 1970. The percentage of persons missing a given item was computed. On the assumption that college freshman are fully competent speakers of English, and thus have a mastery of basic English syntax, any item with a greater than 4.65% error rate was rejected as being an invalid measure of English syntax. The final forms of the LST and LSD were developed from the fifty-six valid items remaining as a result of this process. Thus, forty items on each test were obtained which could be considered to be valid measures of English syntax.

As was noted earlier, 'cloze' procedure items appear to measure the ability to restore texts, an ability which may be independent of linguistic competence. The LSD and LST were, therefore, administered to American fourth-graders in order to obtain some measure of the intrinsic difficulty of the tests themselves. This normative data is presented in Table 3.2 which contains the mean scores and other statistics for each of the tests.

The LSD proved to be only slightly more difficult for the American fourth-graders. Their performance on the tests will be used later as norms for native language competence in computing the number of years of English study required by a Japanese learner to reach that competence.

TABLE 3.2

Means, Standard Deviations, Standard Error of the Mean, and Correlation Coefficients of the Performance of the American Subjects with the Japanese H-Groups.

Group	LST					LSD				
7	N	$\bar{X}^b$	s	$SE_{\bar{x}}$	r	N	$\bar{X}^b$	s	$SE_{\bar{x}}$	r
American	173	.866	.107	.008	----	181	.859	.162	.012	----
HO1	256	.893	.178	.011	.650	253	.805	.179	.011	.401
HO2	220	.727	.223	.015	.612	229	.570	.253	.017	.406
HO3	183	.620	.219	.016	.459	187	.360	.224	.016	.413
HO4	91	.880	.185	.019	.617	229	.784	.189	.013	.352[c]
HO5	196	.841	.199	.014	.571	246	.701	.236	.015	.355[c]
HO6[a]	---	----	----	----	----	251	.837	.155	.010	.455
HO7	244	.881	.173	.011	.650	241	.815	.165	.011	.426
HO8	225	.888	.184	.012	.590	230	.788	.183	.012	.437
HO9[a]	---	----	----	----	----	229	.573	.210	.014	.422
H10	219	.829	.191	.013	.639	218	.678	.222	.015	.337[c]
H11	235	.888	.181	.012	.735	231	.833	.180	.012	.411

a - HO6 and HO9 did not take the LST.
b - Means were computed on the basis of percentage correct.
c - Greater than .025 chance probability. The others are greater than .005 probability.

For the performance of the individual items on the tests, the percentage correct for the American students range in the following way:

LST   55% to 99% with a mean of 87%

LSD   18% to 99% with a mean of 86%

The means for the H-groups range in the following way:

LST   62% to 89% with a grand mean of 83%

LSD   36% to 84% with a grand mean of 70%

It is evident that the tests were more difficult for the Japanese.

The correlations of the performance of the American students with the H-groups range in the following way:

LST   .46 to .73 with a mean of .61

LSD   .34 to .46 with a mean of .40

As could be expected statistically, in general, the higher the correlation, the closer the H-group mean approaches the mean of the English-speaking fourth graders. That there is a considerable difference between the correlations for the LST and LSD may be interpreted in the following way: the LST, by limiting the kind of response required, reduces the degree of freedom of response significantly. This may be seen in the fact that, for the Japanese students, the LST was considerably easier than the LSD--that is, they tended to score consistently higher on the LST than on the LSD. As was shown, this was not the case for the American students.

Since both tests can be considered as valid measures of English syntax, the performance of the American fourth-graders will be considered the INTRINSIC ENGLISH DIFFICULTY of the test items. Since the only commonalities between the American and Japanese groups were the test tasks, approximately 37% of the variance for the LST and 16% for the LSD can be accounted for by this intrinsic English difficulty. The differences between the performance on the tests may then be interpreted as being a result of the fact that the LST tended to direct the Japanese student's attention more directly to the problem of English structure, thus maximizing the effect of intrinsic English difficulty, while the LSD, by demanding a less restricted type of performance, minimized the effect of intrinsic English difficulty.

An important characteristic of a test is its reliability. In tests designed for collecting a sample of English usage for comparison with predictions obtained by contrastive analysis, reliable scores are especially necessary. The aim in the construction of the LST and LSD was to obtain as reliable scores as possible. In this section, evidence of reliability is given, including data on the long-range stability of the LSD scores.

Reliability coefficients were computed for each test, separately for each of the H-groups, V-groups, and the American group. The coefficients are presented in Table 3.3. The average reliability coefficients for H-groups and the V-groups for each of the tests are shown in Table 3.4. These tables indicate that the coefficients are high enough to make useful studies of and draw accurate conclusions about the groups.

The correlation coefficient used was the split-half coefficient, corrected by the Spearman-Brown Prophecy Formula. This type of coefficient is appropriate for the LST and LSD because they are tests in which the speed factor is of little or no importance.

Although the difference between the coefficients for the American students on the LSD and LST (.835 and .822, respectively) is insignificant, they indicate that the test is internally consistent. To obtain a coefficient of .90, the tests would have to be increased to 78 items for the LST and 71 items for the LSD. That the coefficients are somewhat higher than those for the Japanese students suggests that the native speakers are a more heterogeneous group and have, as would be expected, greater syntactic competence.

The difference between the average coefficients for the H-groups on the LSD and LST (.701 and .572, respectively) is significant. That the coefficient for the multiple-choice test, LST, is lower than that for the 'cloze' test, LSD, is surprising. It suggests that multiple-choice testing, a novel method to the Japanese student, may depress the reliability if the examinees are not accustomed to it. This would also explain why the differences between the reliabilities for the native-speakers are insignificant since the average American child is exposed to multiple-choice tests early in his academic career with reading readiness tests, programmed textbooks, etc.

Like the H-group, the difference between the average coefficients on the LSD and LST (.691 and .508, respectively) is significant for the V-groups, also. The reliabilities for the V-groups show an interesting pattern. For the LST, the coefficients progress from .246 for the seventh graders

TABLE 3.3

Means, Standard Deviations, Split-Half Reliability Coefficients, and Standard Errors of Measurement for the LST and LSD by H-Groups

Group	LST					LSD			
	N	$\bar{X}^b$	s	r	SE_M	X^b	s	r	SE_M
H01[c]	251	35.6	2.8	.413	2.15	32.1	3.8	.579	2.47
H02	213	29.0	4.5	.695	2.49	22.6	5.2	.705	2.83
H03	182	24.7	4.6	.620	2.83	14.4	5.9	.800	2.64
H04	91	35.0	3.7	.659	2.16	30.8	5.0	.712	2.69
H05	191	33.5	3.2	.663	1.86	28.0	4.7	.700	2.57
H07	238	35.2	2.8	.554	1.87	32.5	4.0	.666	2.31
H08	221	35.6	2.1	.274	1.79	31.6	4.0	.707	2.16
H10	216	33.1	3.7	.689	2.06	27.1	6.1	.820	2.59
H11	228	35.5	2.2	.517	1.53	33.4	3.3	.619	2.04
V07	180	13.9	3.3	.246	2.86	6.0	3.6	.704	1.96
V08	142	20.3	4.8	.668	2.77	11.6	4.3	.651	2.54
V09	144	28.9	5.1	.799	2.29	21.2	6.5	.860	2.44
V10[c]	251	35.6	2.8	.413	2.15	32.1	3.8	.579	2.47
V11	227	35.6	2.3	.480	1.66	34.7	3.6	.731	1.87
V12	212	36.0	2.4	.414	1.84	34.0	3.3	.642	1.97
American	170	34.7	4.5	.822	1.90	34.6	4.1	.835	1.67

a - Computed by the formula $SE_M = s\sqrt{1-r}$

b - Means were computed on the basis of raw scores.

c - H01 and V10 are the same group.

TABLE 3.4

Average (Mean) Mean Scores, Standard Deviations, Split-Half Reliability Coefficients,[a] and Standard Errors of Measurement[b] of the LST and LSD for the H-groups and V-groups.

Group	LST				LSD			
	Average $\overline{X}$	Average s	Average r	Average SE_M	Average $\overline{X}$	Average s	Average r	Average SE_M
H-group	33.0	3.3	.572	2.16	28.7	4.7	.701	2.57
V-group	28.4	3.5	.508	2.46	23.3	4.2	.691	2.34

a - Since it is not correct procedure to average correlations of coefficients, directly, the mean correlation coefficients for this table were obtained by converting each *r* to Fisher's *z* function, weighting each by its appropriate number of cases, averaging the values, then reconverting. See Quinn McNemar, *Psychological Statistics*. New York: John Wiley and Sons, Inc., 1962, p. 140.

b - Computed by the formula: $SE_{M_{av}} = s_{av}\sqrt{1-r_{av}}$

to .799 for the ninth graders, close to that of the American students. Then, there is a drop to .413, .480, and .414 for the tenth, eleventh, and twelfth grades, respectively. This drop reflects the fact that the junior and senior high schools are different schools, but also shows the tendency of multiple-choice items not to measure accurately at upper levels of ability. The higher correlations for the junior high groups suggest that at the beginning stages of language learning the students are more heterogeneous than at higher levels.

The correlation coefficient is one way of estimating a test's reliability. Another meaningful method is the standard error of measurement which indicates the band of error surrounding any test score. In Table 3.3, the standard errors of measurement for each of the groups, by test, are shown. For the American students, the scores appear to be about equally stable, varying only 1.90 and 1.67 from the true scores on the LST and LSD, respectively. The standard errors of measurement for the Japanese students are higher than this, suggesting that they had to do more guessing. This use of guessing is also reflected in a higher standard error of measurement for the LSD than for the LST among the H-groups. The scores on the LSD also tend to be less stable than those for the LST for the V-groups except for the seventh and eighth grades, where the LST scores are less stable. The tests were very difficult for these groups. Since no correction was made for guessing, they tended to choose any answer for the LST while they omitted the item on the LSD thus increasing the instability of the LST scores, for their group.

One of the important questions of concern when collecting a sample of usage is whether the measurements the test yields are consistent over an appreciable span of time. In part this consistency is a matter of test reliability, in part a question of constancy of the trait measured. Obviously, if a test is unreliable to start with, it cannot be expected to yield similar results on two different administrations even over a short time span. On the other hand, a highly reliable test may not be consistent over a long time span if the trait it measures is unstable or if the students have unequal opportunities for growth in the ability.

Evidence as to the consistency of measurement with the LSD is presented in Table 3.5. No similar evidence is available for the LST. The students comprise the HO2 group. They had taken the LST in October and were retested one month later, in November. The degree of consistency, .568, of the students' performance is satisfactory and suggests that there was much guessing on the part of the students.

TABLE 3.5

Test-Retest Correlation Coefficients, Means, and Standard Deviations on HO2 Group Scores on the LSD

N	r	First Time		Second Time	
		X^a	s	X^a	s
122	.568	22.0	5.0	25.5	4.6

a - Computed on the basis of raw scores.

All in all, the consistency of students' performance on the LST and LSD is such as to permit the use of the data obtained by these instruments for comparison with predictions of errors made by contrastive analyses. Despite different kinds and amounts of practice in the syntactic structures measured by the tests--in school and out--the students generally maintain their relative ranks. This fact speaks well for the stability with which the structures are measured by these tests.

3.8 Intercorrelations of the Tests.

The Investigators needed to know how much confidence they could place in differences between the test scores earned by a particular group on the LST and LSD. This section presents data demonstrating that the LSD and LST have sufficient reliability and independence to permit differential treatment.

Intercorrelations of the Tests. When tests are used in pairs, some characteristics of the tests assume greater significance than when a test is used singly. High reliability is an important and desirable characteristic for any test, although for single proficiency tests reliability is of secondary importance when contrasted with validity. However, when two tests are used together to describe groups so that statements regarding differences among their abilities can be made, the importance of using reliable tests is greatly increased.

The second characteristic essential to tests used in combination is independence from one another. If two tests correlate highly with each other, a person's standing on one usually will not be far from his standing on the other. Consequently, the likelihood of discovering important differences in abilities is seriously diminished for any pair of tests with high intercorrelations, whether this be due to close similarity of content, method of testing, or to other factors.

Table 3.6 presents the intercorrelations of the LST and LSD for each of the groups. Average intercorrelations for the H-groups and the V-groups are shown in Table 3.7. The intercorrelations range in the following way:

H-Group .460 to .765 with a mean of .664

V-Group .292 to .718 with a mean of .563

The mean intercorrelations compare favorably with the coefficient for the American group of .568.

TABLE 3.6

Intercorrelation Coefficients, Means, and Standard Deviations for the LST and LSD by Groups

Group	LST				LSD	
	N	r	$\bar{X}^a$	s	$\bar{X}^a$	s
HO1[b]	251	.561	35.6	2.8	32.1	3.8
HO2	213	.570	29.0	4.5	22.6	5.2
HO3	182	.628	24.7	4.6	14.4	5.9
HO4	91	.765	35.0	3.7	30.8	5.0
HO5	191	.592	33.5	3.2	28.0	4.7
HO7	238	.568	35.2	2.8	32.5	4.0
HO8	220	.471	35.6	2.1	31.6	4.0
H10	215	.687	33.1	3.7	27.1	6.1
H11	228	.460	35.5	2.2	33.4	3.3
VO7	180	.292	13.9	3.3	6.0	3.6
VO8	142	.539	20.3	4.8	11.6	4.3
VO9	145	.718	28.8	5.1	21.2	6.5
V10[b]	251	.561	35.6	2.8	32.1	3.8
V11	227	.453	35.6	2.3	34.7	3.6
V12	212	.490	36.0	2.4	34.0	3.3
American	170	.568	34.7	4.5	34.6	4.1

a - Computed on the basis of raw scores.
b - HO1 and V10 are the same group.

TABLE 3.7

Average (Mean) Intercorrelation Coefficients[a], Means, and Standard Deviations for the LST and LSD for H-Groups and V-Groups

Group		LST		LSD	
	Average r	Average $\overline{X}^b$	Average s	Average $\overline{X}^b$	Average s
H-Group	.664	33.1	4.5	28.3	7.4
V-Group	.563	29.6	9.1	25.0	11.9

a - Since it is not correct procedure to average correlations of coefficients directly, the mean intercorrelation coefficients for this table were obtained by converting each r to Fisher's z function, weighting each by its appropriate number of cases, averaging the values, then reconverting. See Quinn McNemar, Psychological Statistics, New York: John Wiley and Sons, Inc., 1962, p. 140.

b - Computed on the basis of raw scores.

In general, the coefficients demonstrate that the two tests are measuring much of the same thing but in a different way and that they are sufficiently different to warrant the inclusion of the data from both tests in this study.

3.9 Summary of the Chapter.

A multiple-choice test, LST, and a 'cloze' test, LSD, were developed to provide a standardized procedure for collecting a sample of the English usage of Japanese. These tests were found to be sufficiently valid, reliable, and different from one another so that the data obtained by these instruments can be used for comparison with predictions of errors made by selected contrastive analyses.

Eleven groups of Japanese tenth graders were tested in order to determine the degree to which the geographic, dialectal, pedagogical, socio-economic, etc., differences between these groups, designated as the HORIZONTAL VARIABLE, has a significant influence of their English performance, if any. A second group of Japanese junior and senior high school students at six different grade levels were also tested in order to measure the degree to which particular errors are retained in time, designated as the VERTICAL VARIABLE. In the following chapter, the effects of the horizontal and vertical variables on the performance of the Japanese students will be discussed.

A third group of American fourth-graders was tested in order to obtain a measure of the intrinsic difficulty of the tests themselves and to provide normative data for computing the number of years of English study required by a Japanese learner to reach the native speaker's competence. The amount of time required to reach native speaker's competence was designated as TENACITY and will be discussed in the following chapter.

The tests were found to be less difficult and more reliable for these American fourth-graders than for the Japanese students. That the tests were difficult, especially for the seventh and eighth graders, forced the Japanese students to do more guessing and thus suppressed the reliability coefficients. The high reliability coefficients for the American students suggest that they are a more heterogeneous group in syntactic competence than non-native learners of English. That this is so is perhaps shown by the differences in the processes of first and second language acquisition where in acquiring his native language a person is exposed to a random sample of that language, thus producing greater heterogeneity among native speakers, while in acquiring a second language the learner is exposed to a selected sample of the language, thus producing greater homogeneity among non-native speakers. This homogeneity of error production among the Japanese students was also observed in examining the data for the effects of the horizontal and vertical variables and will be discussed in greater detail in the following chapter.

CHAPTER IV. THE PATTERN OF JAPANESE ERRORS

4.0 In the previous chapter, the method used in collecting a sample of the usage of Japanese students was described. In this chapter, the pattern of errors of the eleven groups of Japanese tenth-graders, i.e., the H-groups, and of the six groups of Japanese junior and senior high school students, i.e., the V-groups, will be examined. From this examination, patterns of errors will be derived for later comparison with the predictions of errors made by selected contrastive analyses.

4.1 The Horizontal Variable.

The Horizontal Variable was defined earlier as the degree to which the differences -- geographic, dialectal, socio-economic, pedagogical, etc. -- between groups of learners at the same educational level would influence their learning of a foreign language. If this variable should affect the students' performance erratically, then contrastive analyses could not be generalized from one dialect to another. If, however, the variable should have little effect on the students' performance, then contrastive analyses could be generalized from a comparison of a single dialect with the target language.

The first test for the Horizontal Variable was to rank the mean scores for each of the H-groups in order from lowest to highest. This ranking for the LST and LSD is shown in Table 4.1. By correcting the means by the standard error of the mean, on both tests, the eleven schools were found to fall into four groups:

a. Group I, consisting of H03, a prefectural technical school in the West Kanto dialect area;

b. Group II, consisting of H02 and H09, prefectural urban and rural schools in the West Kanto and Kinki dialect areas, respectively;

c. Group III, consisting of H10 and H05, a private rural school in the East Kanto dialect area, and a prefectural urban school in the West Kanto dialect area;

d. Group IV, consisting of H04, H07, H08, H01, H11, and H06, four prefectural urban schools, one prefectural rural school, and one private urban school in the West Kanto, Chugoku, and Kinki dialect areas.

The rank orders for the LST and LSD of performance of the schools was tested by the Spearman rank-difference correlation and the resulting coefficient was .867, with a greater than

TABLE 4.1

Rank Order, Means, Standard Deviations, and Standard Error of the Mean on the LST and LSD for the H-Groups.

LST						LSD					
Rank	Group	N	$\bar{X}^b$	s	SE_x	Rank	Group	N	$\bar{X}^b$	s	SE_x
1.	H03	183	.620	.22	.02	1.	H03	187	.360	.22	.02
2.	H02	220	.727	.22	.02	2.	H02	229	.570	.25	.02
	---	---	----	---	---	3.	H09^c	229	.573	.21	.01
3.	H10	219	.829	.19	.01	4.	H10	218	.678	.22	.02
4.	H05	196	.841	.20	.01	5.	H05	246	.701	.24	.02
5.	H04	91	.880	.18	.02	6.	H04	229	.784	.19	.01
6.	H07	244	.881	.17	.01	7.	H08	230	.788	.18	.01
7.	H11	235	.887	.18	.02	8.	H01	253	.805	.18	.01
8.	H08	225	.888	.18	.01	9.	H07	241	.815	.17	.01
9.	H01	256	.893	.18	.01	10.	H11	231	.833	.18	.01
	---	---	----	---	---	11.	H06^c	251	.837	.16	.01

a - Computed by the formula: $SE_x = \frac{s}{\sqrt{N-1}}$

b - Means computed on the basis of percentage correct.

c - H09 and H06 did not take the LST.

.01 chance probability. This would seem to indicate that the LST and LSD work equally well in measuring the relative proficiency of the eleven groups, i.e., the raw scores in the groups are parallel, but not the same.

Even so, there are internal differences between the four groups which may be interpreted as the operation of the Horizontal Variable. What the specific factors operating in the variable are is unknown, but would appear to be neither geographic nor dialectal since urban and rural schools and different dialect areas are represented in Groups II, III, and IV. Socio-economic factors, however, may be operating in the variable since the students in Groups I and II are generally from a lower economic class than those in Groups III and IV. This socio-economic factor may also be reflected in pedagogy, but what the specific factors are was beyond the scope of this study. The purpose of this study was merely to find out whether a Horizontal Variable does in fact operate in foreign language learning and the examination of the data presented would seem to indicate that it is indeed operating.

The second test on the Horizontal Variable was to determine its effect, if any, on the performance of Japanese students. This was done by correlating the response patterns of the various H-groups. For each of the forty items on the LSD and LST, the percentage of grammatical responses, i.e. difficulty, was correlated with that of all of the other H-groups. These correlations are shown in Tables 4.2 and 4.3 for the LST and LSD, respectively. The correlation of profile of performance is illustrated graphically in Figure 4.1 in which the response patterns on the LST for H01 and H08 are shown, and for which the correlation coefficient was .977.

TABLE 4.2

Intercorrelations of the LST Profiles of Performance for the H-Groups.

	HT01	HT02	HT03	HT04	HT05	HT07	HT08	HT10	HT11
HT01	1.000								
HT02	.794	1.000							
HT03	.608	.908	1.000						
HT04	.972	.814	.674	1.000					
HT05	.941	.855	.696	.921	1.000				
HT07	.961	.749	.606	.970	.909	1.000			
HT08	.977	.763	.603	.972	.935	.982	1.000		
HT10	.901	.939	.800	.879	.920	.839	.874	1.000	
HT11	.954	.794	.624	.952	.862	.960	.946	.873	1.000

p = .005

Fig. 4.1

Performance profiles of the HO1 and HO3 groups on the LST

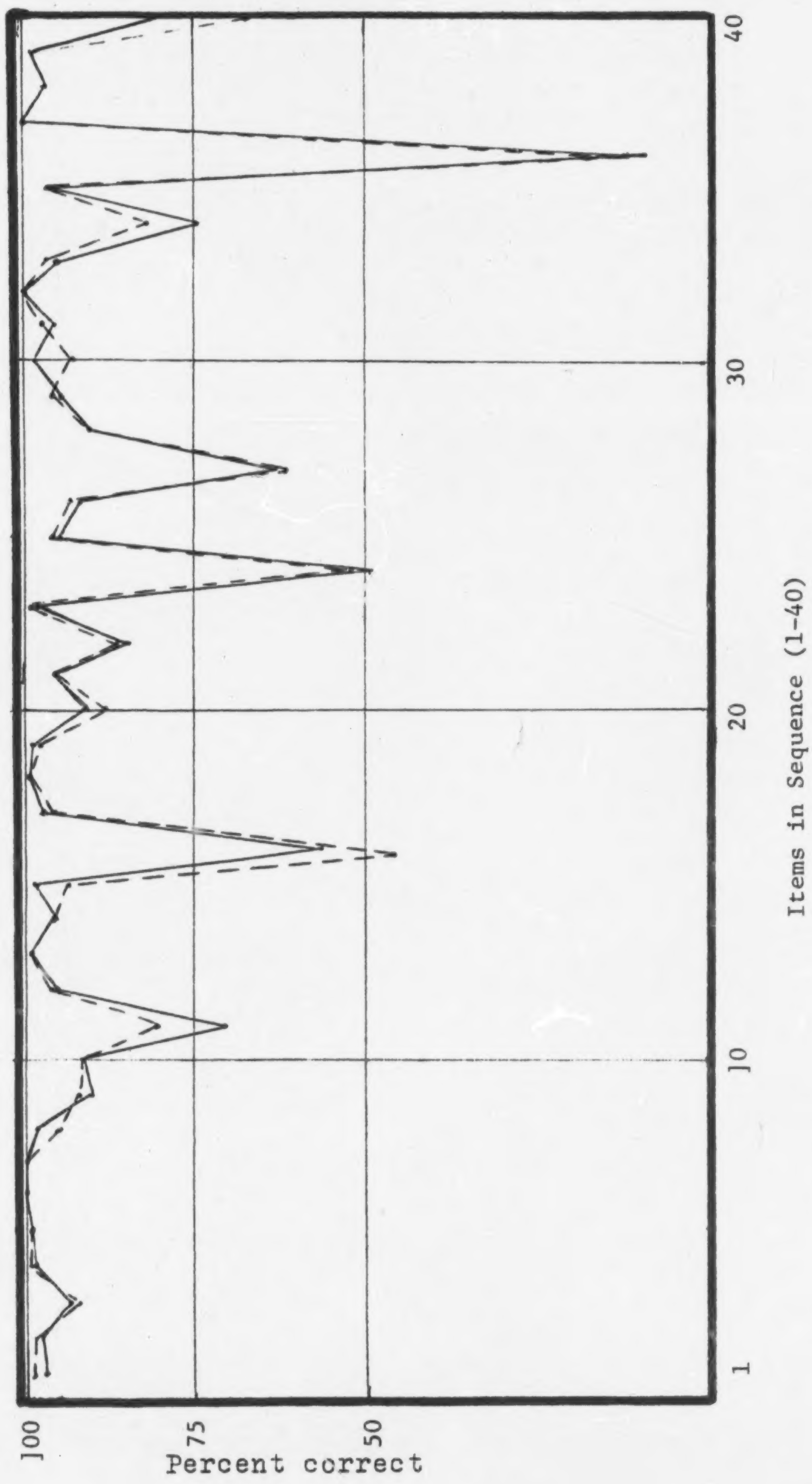


TABLE 4.3

Intercorrelations of the LSD Profiles of Performance for the H-Groups.

	HD01	HD02	HD03	HD04	HD05	HD06	HD07	HD08	HD09	HD10	HD11
HD01	1.000										
HD02	.714	1.000									
HD03	.678	.905	1.000								
HD04	.852	.773	.685	1.000							
HD05	.826	.889	.805	.892	1.000						
HD06	.916	.688	.661	.895	.782	1.000					
HD07	.928	.792	.735	.912	.880	.944	1.000				
HD08	.859	.759	.731	.788	.866	.815	.897	1.000			
HD09	.781	.896	.896	.813	.845	.785	.806	.772	1.000		
HD10	.845	.917	.851	.836	.935	.788	.867	.898	.885	1.000	
HD11	.877	.676	.641	.843	.812	.871	.859	.841	.782	.828	1.000

p = .005

The correlations of the profiles of performance for the two tests range in the following way:

LST: .603 to .982 with a mean of .894
LSD: .641 to .944 with a mean of .840

It is perhaps not surprising that the lowest correlations tend to be scored by those having the lowest mean scores, H03 (rank 1, Group I) and then H02 (rank 2, Group II). Yet the lowest groups tend to correlate highly with one another as is shown in Table 4.4.

TABLE 4.4

Intercorrelations of the LST and LSD Profiles of Performance for the Three Lowest H-Groups: H03, H02, H09.

	LST		LSD		
	H03	H02	H03	H02	H09[a]
H03	1.000		1.000		
H02	.908	1.000	.905	1.000	
H09[a]	----	----	.896	.896	1.000

p = .005
a - H09 did not take the LST.

Their relatively poor correlations with the higher scoring groups may be a function of raw score shifts and not internal differences. The highest correlations, as can be seen, generally occur within groups ranked closest to each other on the basis of mean scores.

An estimate of the amount of variance accounted for by the commonalities of the H-groups is 79% for the LST and 71% for the LSD, or very roughly 75% for both. These commonalities of the Japanese tenth-graders are as follows:

a. Common language and cultural system;
b. Common educational system;
c. Common level of English and age;
d. Common test of English.

It may be hypothesized that the remaining 25% is accounted for by differences, such as individual differences among students, local variations discussed above, and pedagogical differences from school to school. The degree to which a contrastive analysis takes these commonalities and differences into account would seem to indicate probable success in predicting the errors students will make. It was hypothesized that the Horizontal Variable would affect target language production erractically. However, the high degree of correlation between the profiles of performance for the eleven H-groups suggests that whatever effect the variable has, it is fairly constant for all learners of a given target language.

4.2 The H-Pattern.

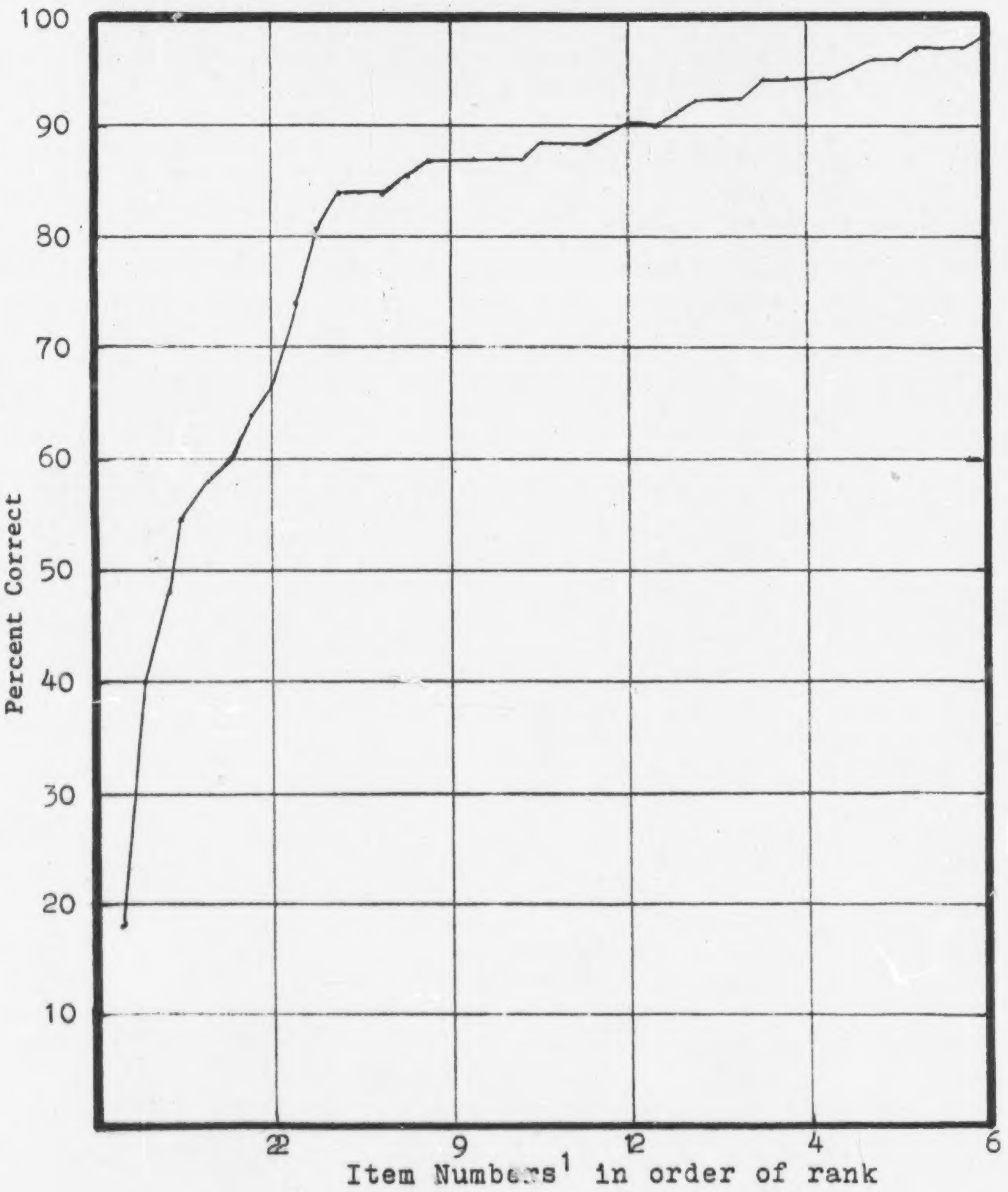
Having found a high correlation among the profiles of performance of the H-groups, the specific pattern of response was next defined. In this section, the nature of that pattern will be discussed.

For each of the forty items on the LST and LSD for each the groups, the percentage of grammatical responses, i.e., difficulty, were ranked in order from least to most difficult. The degree of relation among the nine rankings of the forty items on the LST and the eleven rankings on the LSD was tested by a Kendall Coefficient of Concordance, W (Siegel, Formula 9.15, p. 231). The resulting coefficients corrected for tied values were .804 for the LST and .859 for the LSD both with a chi-square probability of greater than .001. (The estimate of the average value of the Spearman rank correlation was .780 for the LST and .845 for the LSD. See Siegel, Formula 9.14, p. 229). The high correlations are surprising because it suggests that whatever effect the Horizontal Variable has on the production of errors, it applies more or less equally to different structures. Furthermore, it suggests that the output of a contrastive analysis can indeed be generalized for all speakers of a language even though only one dialect in the source language is contrasted with a single dialect of the target language.

Since the coefficients of concordance are so high, the mean percentage of grammatical responses for all H-groups defined as P can be used as the Horizontal Pattern of the English usage of Japanese students. This P will later be correlated with predictions of difficulty and predictions of degree of difficulty to obtain indices of the qualitative power of a given contrastive analysis to predict errors. The Horizontal Pattern, P, for the LST and LSD in rank order order from least to most difficult is given in Table 4.5. The Horizontal Patterns are displayed graphically in Figures 4.2 and 4.3 for the LST and LSD, respectively.

Fig. 4.2

Horizontal Pattern of the H-Groups on the LST

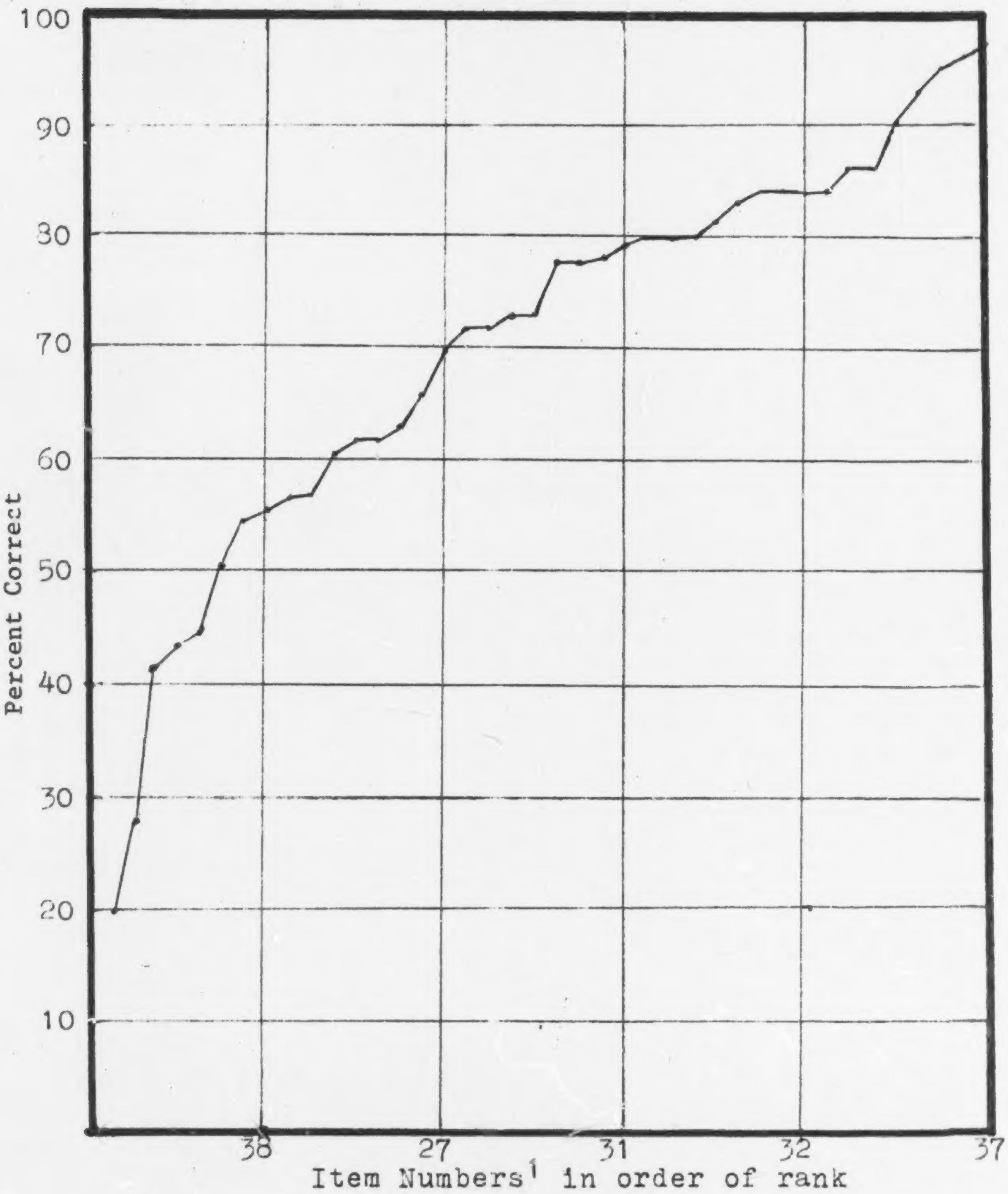


---

[1] Every eighth item is listed in the Figure 4.2 For full list, see Table 4.5.

Fig. 4.3

Horizontal Pattern of the H-Groups on the LSD



---

[1] Every eighth item is listed in Figure 4.3. For the full list, see table 4.5.

47-48

TABLE 4.5

The Horizontal Pattern, P, of the Errors of Japanese Students in Rank Order from Least to Most Difficult for the LST and LSD

	LST	
Rank	Description[a]	Mean Percent Correct
1.	Aux + Ving: are (6)	.985
2.	Question: did (32)	.977
3.	Negative: not (1)	.975
4.	Pronoun: him (7)	.974
5.	Interrogative: what (37)	.970
6.	Negative: can't (5)	.969
7.	V + ed: began (18)	.952
8.	V + ed: was (2)	.947
9.	Count N + sg: house (4)	.946
10.	Negative: don't (38)	.944
11.	V + Ø: make (13)	.942
12.	Relative: when (33)	.929
13.	Interrogative: when (23)	.926
14.	Aux + Ven: is (19)	.922
15.	Negative: not (35)	.918
16.	Question: may (17)	.910
17.	Question: would (12)	.903
18.	Pre-determiner: one (26)	.892
19.	Adjective: personal (31)	.890
20.	V + en: become (39)	.888
21.	Pronoun: her (21)	.887
22.	Question: did (14)	.877
23.	V + Ving: saw (30)	.875
24.	V + ing: making (8)	.873
25.	V + ed: had (9)	.873
26.	Adjective: important (10)	.872
27.	Function Noun: any (3)	.860
28.	Aux + Ven: had (25)	.849
29.	Relative: which (29)	.846
30.	Aux + VØ: must (15)	.841
31.	Adjective: little (20)	.820
32.	V + to V: want (28)	.732
33.	Mass Noun: information (22)	.678
34.	Determiner: the (40)	.645
35.	Aux + to V: used (11)	.603
36.	Interrogative: how (34)	.589
37.	Relative: that (27)	.561
38.	Count Noun + pl: years (16)	.465
39.	Determiner: each (24)	.409
40.	Aux + Ven: was (36)	.182

Grand Mean: .827
Grand Standard Deviation: .21

TABLE 4.5 (continued)

Rank	LSD Description[a]	Mean Percent Correct
1.	Interrogative: what (37)	.970
2.	Pronoun: him (7)	.962
3.	Aux + Ving: are (6)	.949
4.	Aux + VØ: must (15)	.927
5.	Question: would (12)	.899
6.	Relative: which (29)	.865
7.	Aux + Ven: is (19)	.857
8.	Pronoun: her (21)	.842
9.	Question: did (32)	.840
10.	Question: did (14)	.840
11.	Aux + Ven: had (25)	.839
12.	Negative: not (1)	.835
13.	Negative: can't (5)	.809
14.	V + ed: began (18)	.805
15.	Interrogative: when (23)	.800
16.	Determiner: then (40)	.797
17.	Adjective: personal (31)	.788
18.	Count Noun-sg: house (4)	.785
19.	Determiner: each (24)	.784
20.	V + ing: making (8)	.782
21.	Question: may (17)	.732
22.	V-ed: was (2)	.729
23.	Pre-determiner: one (26)	.724
24.	V + en: become (39)	.722
25.	Relative: that (27)	.695
26.	Aux + Ven: was (36)	.659
27.	V + to V: want (28)	.628
28.	Adjective: important (10)	.624
29.	V + Ving: saw (30)	.616
30.	Aux + to V: used (11)	.605
31.	Adjective: little (20)	.574
32.	Negative: not (35)	.568
33.	Negative: didn't (38)	.561
34.	Interrogative: how (34)	.547
35.	V + ed: had (9)	.518
36.	Count Noun + pl: years (16)	.448
37.	Relatives: when (33)	.441
38.	Function Noun: any (3)	.422
39.	V + Ø: make (13)	.276
40.	Mass Noun: information (22)	.190

Grand Mean: .706
Grand Standard Deviation: .25

a - The descriptions are given in the following way: syntactic pattern tested; correct answer for LST and possible filler for the blank in LSD; number of item in parentheses. Underlined entries indicate slot in pattern to be filled.

The rankings on the Horizontal Pattern of the LSD were tested against those for the LST by the Spearman rank-difference correlation. The resulting coefficient was .43 with a chance probability of greater than .001 which suggests that the predictions of a given contrastive analysis may correlate better with the performance on the LST than on the LSD or vice versa, but it will probably not correlate highly on both.

In order to test the gross capacity of a contrastive analysis to predict difficulty, a variable E was derived from the mean percentage of grammatical responses, P, to represent gross occurrence of error. E is a two-valued variable having the value 0 if the mean value of an item is above the grand mean of all items on the test, and the value 1 otherwise. For the LSD, the 24 items ranked 1 to 24 on Table 4.5 had a value of 0 since they were above the grand mean for the LSD of .706. For the LST, the 30 items ranked 1 to 30 on Table 4.5 had a value of 0 since they were above the grand mean for the LST of .827. These values will later be correlated with predictions of difficulty to obtain an index of gross capacity to predict errors. However, since E's having a value of 0 were to be correlated with the lowest levels of predicted difficulty which tended to rate less than ten items as representing 'no difficulty' for either test, it was felt that the disparity in the input ratings might artificially lower the correlations. For this reason, several other "E's" were derived.

The variable E_1 was calculated from the mean percentage of correct Japanese responses for an item and the equivalent English percentage for the same item. If the Japanese percentage for the item was higher than the English percentage correct, then the item was given as E_1 value of 0, otherwise 1. In effect, the variable E_1 represents a gross occurrence of performance relative to the English performance. On the LSD, five items were valued 0, and on the LST, 21 items.

The variable E_2 was calculated directly from mean percentage correct for each item. If the Japanese mean percentage correct for item was greater than .950, then it was given an E_2 value of 0, otherwise 1. As can be seen on Table 4.5, two items on the LSD were valued 0, and seven items on the LST.

The variable E_3 was calculated like E_2 from the mean percentage correct directly. If the mean percentage were greater than .900, then it received an E_3 value of 0, otherwise 1. As can be seen on Table 4.5, four items on the LSD were valued 0 and seventeen items on the LST. Both E_2 and E_3 rate difficulty on an absolute basis of 95% or 90% correct, respectively.

4.3 Vertical Variable.

The Vertical Variable was earlier defined as differences in error production caused by different amounts of time spent in learning the target language. It was hypothesized that this variable would not affect the performance of Japanese students arbitrarily, but would decrease in effect through time. Contrastive analyses often make their predictions of difficulty based on the assumption that the problems predicted are those which a person who has had no exposure to the target language will encounter in learning that language. From this, it is theorized by Lado (1964, p. 52) and others that language teaching should concentrate on these problems and when they have been mastered the language will have been mastered. However, it may be that some problems in the target language can never be mastered and would characterize a particular non-native dialect of English, for example, "Japanese" English. Such problems which could not be mastered could be considered tenacious. The degree to which a contrastive analysis takes this tenacity of errors into account would seem to indicate probable success in predicting levels of difficulty students will face in learning the target language.

In order to test the degree to which the Vertical Variable affected the English performance of Japanese students, the means of the six V-groups were compared. The means and other statistical data for the V-groups are given in Table 4.6. As was hypothesized, the Vertical Variable is not arbitrary in nature and does decrease in effect through time.

TABLE 4.6

Means, Standard Deviations, and Standard Errors of the Mean[a] on the LST and LSD for the V-groups

Grade	LST				LSD			
	N	X^b	s	$SE_{\bar{X}}$	N	X^b	s	$SE_{\bar{X}}$
7	186	.347	.20	.01	193	.146	.16	.01
8	145	.507	.24	.02	150	.289	.25	.02
9	145	.723	.22	.02	146	.530	.25	.02
10[c]	256	.893	.18	.01	253	.805	.18	.01
11	231	.890	.19	.01	243	.866	.13	.01
12	214	.899	.17	.01	236	.854	.15	.01

a - Computed by the formula: $SE_{\bar{X}} = \frac{s}{\sqrt{N-1}}$

b - Means computed on the basis of percentage correct.

c - Grade 10 is the same group as H10.

The effect of the Vertical Variable is shown graphically in Figure 4.4 for the LST and Figure 4.5 for the LSD. As can be observed in the graphs, after about three years of language study, the average Japanese student tested performed as well or better than the average American fourth-grader tested.

The Vertical Variable seems to reflect the influence of the intrinsic English difficulty of the tests. This intrinsic difficulty does not appear to be constant, and seems to depend to a great extent on the amount of English that the learner already knows. To test the effects of this intrinsic difficulty, a Spearman rank-difference correlation of the rank order of the percentages of difficulty for each of the forty items on the LST and LSD for the American students and the V-groups was obtained since the only commonality involved would be that of the English test. The resulting coefficients and the amount of variance accounted for are given in Table 4.7.

TABLE 4.7

Rank-Difference Correlation Coefficients and Variance for the Rankings of Percentage of Difficulty on the LST and LSD for the American and V-Groups

—	LST		LSD	
Grade	rho	V	rho	V
7	$.16^{a}$	3%	$.23^{a}$	5%
8	$.62^{e}$	38%	$.31^{b}$	10%
9	$.51^{e}$	26%	$.38^{c}$	14%
10	$.60^{e}$	36%	$.36^{c}$	13%
11	$.61^{e}$	37%	$.40^{c}$	16%
12	$.59^{e}$	35%	$.42^{d}$	18%

a - p=.10
b - p=.05
c - p=.01
d - p=.005
e - p=.001

As can be seen, intrinsic English difficulty appears to play little role initially when the language learner is characterized by a gross ignorance of what English is about, but plays an increasing role as he learns more and more English. Although the data is limited, it suggests that by the ninth or tenth grade, a Japanese student approaches a point where intrinsic English difficulty tends to level off, at about 15% for the LSD and about 35% for the LST, or at least to increase much more slowly.

Fig. 4.4

Vertical Pattern of the V-Groups on the LST

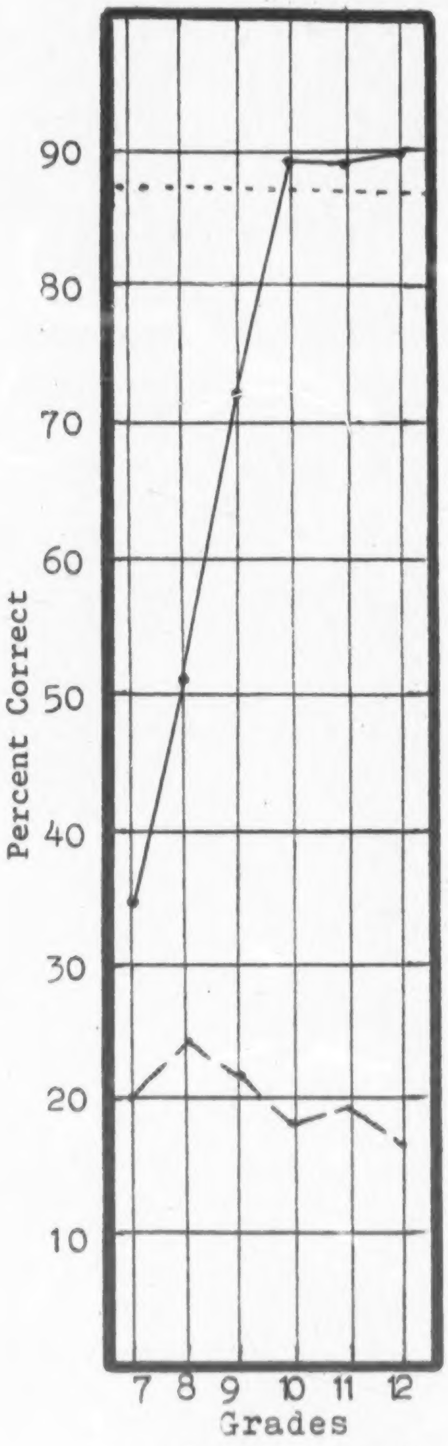
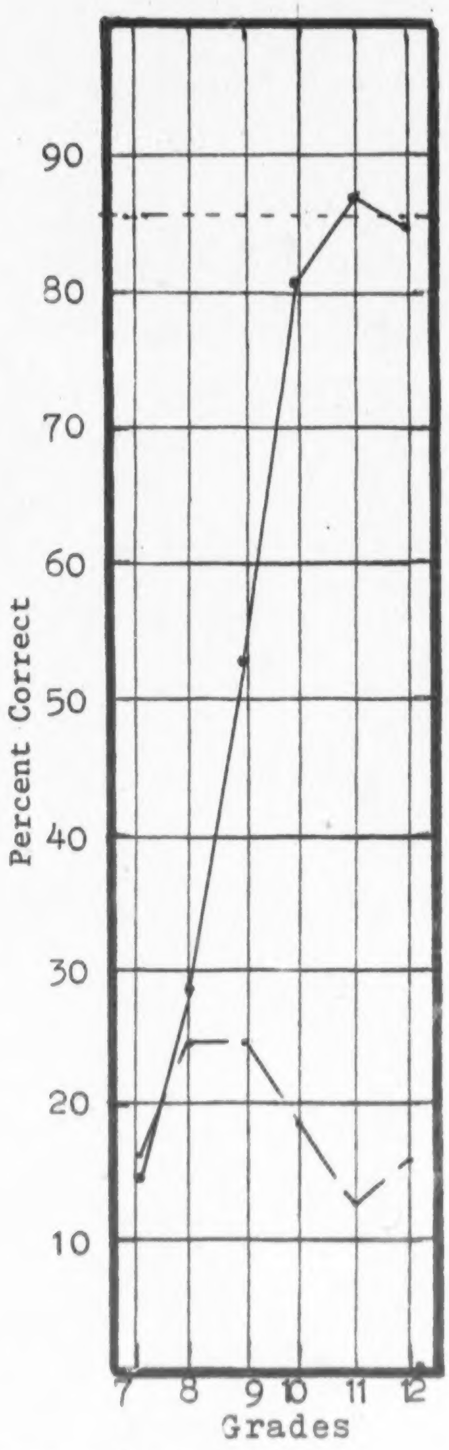


Fig. 4.5

Vertical Pattern of the V-Groups on the LSD



---- A-Groups(American) mean percentage correct
—— V-Groups
— — V-Groups Standard Deviations

By doing a rank-difference correlation of the rank order of the percentages of difficulty for each of the forty items on the LST and LSD for the H-groups and V-groups, suggestive information about the role played by the Horizontal-Vertical commonalities was elicited. These commonalities are:

a. Common native language and culture.
b. Common educational system at the highest level.
c. Common tests of English.

The Horizontal-Vertical correlations and amounts of variance accounted for is reported in Table 4.8.

TABLE 4.8

Rank-Difference Correlation Coefficients and Variance for the Rankings of Percentage of Difficulty on the LST and LSD for the H-Groups and V-Groups

	LST		LSD	
Grade	rho	V	rho	V
7	.28[a]	8%	.48	23%
8	.73	63%	.52	27%
9	.82	68%	.83	69%
10	.83	69%	.91	83%
11	.88	77%	.87	76%
12	.88	77%	.88	77%

a - p=.05. For all other correlations, p=.001

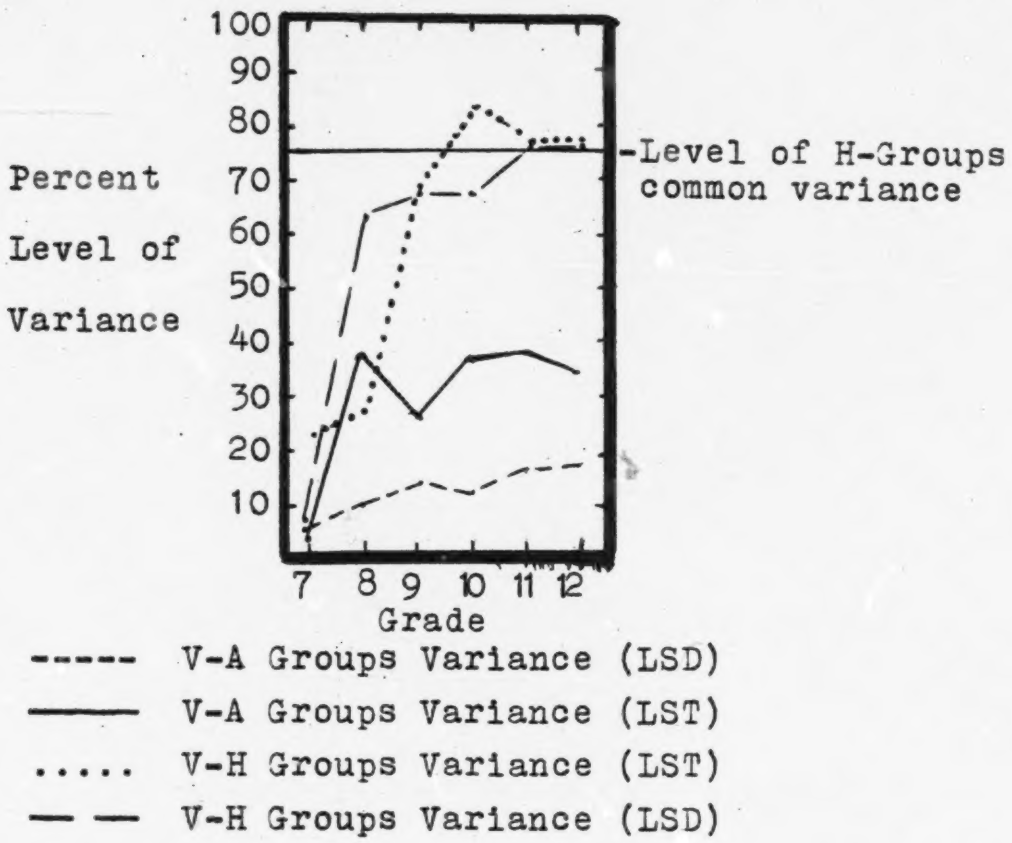
For purposes of comparison, the Vertical Horizontal variances and the American-Vertical variances are plotted graphically in Figure 4.4. The interpretation of this figure is necessarily vague, since variances cannot be confidently subtracted from one another. However, it suggests that common native language and culture and common educational system play a significant role in performance similarity, more or less to the degree that the Vertical-Horizontal correlations rise above the American-Vertical correlations. That is, given the data at hand, from the ninth or tenth grade on, between 40% and 60% of the variance may be attributable to the fact that the students all have a common language and culture, and they are all being educated in Japan.

In summary, the roles played by the commonalities between the Vertical and Horizontal groups seem to be for the ninth, tenth, eleventh, and twelfth graders the following:

2

Fig. 4.6

V-A and V-H Variance Characteristics



a. Common language and culture, and	
b. Common educational system, together	40-60%
c. Common test of English, i.e., intrinsic difficulty of the instrument	15-35%
d. Common level of English proficiency	No detectable influence for the tenth grade, no data otherwise.

A great deal more data would be required to test these hypothetical values more deeply. No data at all is available to define specifically the role of language and culture in isolation. This role of language and culture would seem to be a key factor in assessing the potential value of the contrastive analysis hypothesis which in its strong form attempts to predict errors on the basis of linguistic descriptions and in its weak form to explain errors. (Wardhaugh, 1970)

4.4 The Vertical Pattern.

Having observed the operation of the Vertical Variable, the specific pattern of responses was next defined. In this section, the nature of this pattern will be discussed.

For each of the forty items on the LST and LSD, the percentage of grammatical responses, i.e., difficulty, were ranked in order from least to most difficult for each of the six V-groups. The degree of relation among the six rankings was tested by a Kendall Coefficient of Concordance, W. (Siegel, Formula 9.15, p. 231). The resulting coefficients corrected for tied values was .72 for the LST and .70 for the LSD, both with a chi-square probability of greater than .001. (The estimate of the average Spearman rank correlation is .67 for the LST and .64 for the LSD. See Siegel, Formula 9.14, p. 229). This rather high correlation is surprising since it suggests that items that are initially relatively difficult tend to remain relatively difficult over a period of six years of language study. Furthermore, it suggests that the output of a contrastive analysis, if it can predict relative problems for the language learner, could do so almost as adequately for the sixth year student as for the first.

The percentage of grammatical responses for each item on the two tests arranged in order from the seventh grade to the twelfth grade was compared with the pattern of responses of the American fourth-graders in order to obtain an estimate of the number of years of English study it would take for the average Japanese student to reach the level of performance of the average English-speaking fourth grader. This estimate was defined as TENACITY, T. The formula for deriving T together with the Tenacity levels for each of the items on the LST and LSD are given in Table 4.9. In some cases, the Japanese percentage of difficulty at the twelfth grade, i.e., after six years of English study, was below that of the American students. In such cases, T was derived by projecting the number of years beyond twelfth grade, it would require for the Japanese student to reach . the level of the American students. The formula for deriving this projected estimate of T is also given in Table 4.9.

The rank orders of Tenacity on the LST and LSD were tested for rank-difference by the Spearman rank-difference correlation. The resulting coefficient was .56 with a chance probability of greater than .001 indicating that the orderings were fairly similar. This factor T will be correlated with predictions of difficulty obtained by contrastive analysis in order to obtain an index of a given analysis' sensitivity to difficulty over time. A given analysis may correlate better with the Tenacity factor on the LST than with the LSD or vice versa, but not on both because the correlation of ranks between the tests is not high.

4.5 Summary of the Chapter.

It is clear that a Horizontal Variable, the effects of geographic, socio-economic, dialectal, pedagogical, etc., factors on language learning does operate in second language learning and whatever effect it has, it is fairly constant for all learners of a given target language. The variable also appears to apply more or less equally to different structures. This suggests that the output of a contrastive analysis can be generalized for all speakers of a given language even though only a single dialect of the source language is contrasted with a single dialect in the target language. A Horizontal Pattern, P, was defined for the forty items on the LST and LSD in terms of the mean percentages of grammatical responses of the H-groups. This Horizontal Pattern will later be correlated with predictions of difficulty from selected contrastive analyses to obtain an index of the qualitative power of an analysis to predict errors. Another variable, E, having two values, 0 and 1, was also derived from the Horizontal Pattern in four forms and will later be correlated with predictions of difficulty to obtain an index of the gross capacity of a contrastive analysis to predict errors.

It is also clear that a Vertical Variable, the effect of time on language learning, operates in second language learning and whatever effect it has, it is not arbitrary and decreases in effect through time. This variable is also reflected in the intrinsic English difficulty of the test

TABLE 4.9

The Tenacity Factor, T, in Rank Order from Least to Most Tenacious for the LST and LSD

	LST	
Rank	Description[a]	T[b]
1.	Question: did (32)	1.83
2.	Interrogative: what (37)	1.96
3.	Aux + Ving: are (6)	2.50
4.	Negative: not (1)	2.65
5.	Adjective: important (10)	2.78
6.	Aux + Ven: was (36)	2.82
7.	Relative: which (29)	2.83
8.	Relative: when (33)	2.93
9.	Pre-determiner: one (26)	2.95
10.	Negative: can't (5)	3.09
11.	V + ed: began (18)	3.22
12.	Adjective: long (31)	3.40
13.	V + Ving: saw (30)	3.42
14.	V + ed: had (9)	3.43
15.	Interrogative: when (23)	3.45
16.	Aux + VØ: must (15)	3.47
17.	V + en: become (39)	3.51
18.	Negative: didn't (38)	3.57
19.	Pronoun: her (7)	3.58
20.	Mass Noun: information (22)	3.59
21.	Question: did (14)	3.61
22.	Aux + Ven: is (19)	3.65
23.	V-ed: was (2)	3.69
24.	Relative: that (27)	3.73
25.	Aux + to V: used (11)	3.78
26.	Aux + Ven: has (25)	3.79
27.	Count N-sg: house (4)	3.83
28.	Negative: not (35)	3.84
29.	V + ing: making (8)	3.88
30.	V + Ø: make (13)	3.89
31.	Question: may (17)	3.93
32.	V + to V: want (28)	3.95
33.	Pronoun: her (21)	3.95
34.	Question: would (12)	4.00
35.	Function Noun: any (3)	4.20
36.	Determiner: the (40)	5.00
37.	Adjective: little (20)	6.17[c]
38.	Interrogative: how (34)	7.46[c]
39.	Determiner: each (24)	9.16[c]
40.	Count Noun-pl: years (16)	10.[illegible]9[c]

TABLE 4.9 (continued)

	LSD	
Rank	Description[a]	T[b]
1.	Adjective: important (10)	1.40
2.	Interrogative: what (37)	2.73
3.	Aux + Ving: are (6)	3.00
4.	Relative: which (29)	3.33
5.	Aux + Ven: is (19)	3.56
6.	V + en: become (39)	3.57
7.	Interrogative: when (23)	3.58
8.	V + ed: began (18)	3.62
9.	Aux + VØ: must (15)	3.80
10.	Negative: can't (5)	3.87
11.	Relative: that (27)	3.88
12.	Question: would (12)	3.89
13.	Question: did (32)	3.92
14.	Aux + to V: used (11)	3.98
15.	Pronoun: him (7)	4.00
16.	Mass Noun: information (22)	4.33
17.	Question: may (17)	4.36
18.	V + Ving: saw (30)	4.50
19.	Aux + Ven: was (36)	4.57
20.	Negative: not (1)	4.78
21.	Count N-sg: house (4)	5.00
22.	V-ing: making (8)	5.00
23.	Pre-determiner: one (26)	5.00
24.	Negative: didn't (38)	5.60
25.	Determiner: the (40)	5.67
26.	Negative: not (35)	5.91[c]
27.	Aux + Ven: has (25)	6.06[c]
28.	Adjective: long (31)	6.07[c]
29.	Pronoun: her (21)	6.23[c]
30.	Count N-pl: years (16)	6.30[c]
31.	Determiner: each (24)	6.46[c]
32.	V-ed: was (2)	6.81[c]
33.	Question: did (14)	6.82[c]
34.	V + to V: want (28)	7.48[c]
35.	Adjective: little (20)	7.57[c]
36.	Relative: when (33)	7.60[c]
37.	V + Ø: make (13)	7.62[c]
38.	V + ed: had (9)	7.83[c]
39.	Function Noun: any (3)	7.97[c]
40.	Interrogative: how (34)	12.00[c]

a - The descriptions are given in the following way: syntactic pattern tested; correct answer for the LST and possible filler for the blank in LSD; number of item in parentheses. Underlined entries indicate slot in pattern to be filled.

TABLE 4.9 (continued)

b - The formula for deriving T is as follows: $T = Y + \frac{E - lcj}{hcj - lcj}$

where:

T = estimate of Tenacity;
Y = year, 1-6, where the percentage of difficulty is just below that of the American students;
E = percentage of difficulty of American students for a given item;
lcj = percentage of difficulty of Japanese students just below E;
hcj = percentage of difficulty of Japanese students just above E.

c - The formula for deriving these estimates of T beyond the sixth year of study is as follows:

$$*T = 6 + \frac{E - hj}{rg}$$

where:

*T = projected estimate of Tenacity;
E = percentage of difficulty of American students;
hj = percentage of difficulty of Japanese twelfth graders;
rg = the average rate of growth for the Japanese students.

The average rate of growth was obtained by the following formula:

$$rg = \frac{\left(\frac{J^7 - J^8}{2}\right) + \left(\frac{J^{11} - J^{12}}{2}\right)}{5}$$

where:

rg = average rate of growth;
J^7 = percentage of difficulty for the Japanese seventh graders;
J^8 = percentage of difficulty for the Japanese eighth graders;
J^{11} = percentage of difficulty for the Japanese eleventh graders;
J^{12} = percentage of difficulty for the Japanese twelfth graders.

where it accounts for 15 to 30% of the variance after two or three years of language study--the point at which the performance of the Japanese students on the tests matched that of the American fourth-graders. At this point, common language and culture and a common educational system appear to be more important factors in error production, accounting for 40 to 60% of the variance. Although the Vertical Variable decreases in effect through time, items which are difficulty initially tend to remain so for a period of six years. This suggests that the predictions made by a contrastive analysis would be as adequate for the first year as for the sixth. An estimate of the number of years of English study required for a Japanese student to reach the level of English performance was defined as Tenacity, T. This variable will be correlated later with predictions of difficult to obtain an index of the sensitivity of the output of a given contrastive analysis to error production over time.

In the following chapter, four different models of contrastive analyses of Japanese and English will be discussed and a set of predictions will be derived for the sentences which appeared as test items in the LSD and LST. These predictions will later be correlated with the P, E, and T variables discussed in this chapter.

CHAPTER V. THE CONTRASTIVE ANALYSES

5.0 This chapter concerns the predictions that contrastive analyses make about the difficulty that Japanese speakers will have on each sentence of the Language Sampling Device (LSD). The first part of this chapter deals with the reasons for selecting the four contrastive analyses that were ultimately investigated (5.1), followed by a discussion of the variables that each analysis manipulates in formulating predictions of difficulty (5.2). The next two sections deal with the establishment of a corpus of Japanese sentences with which the English LSD sentences could be compared (5.3), and the descriptive analysis of both the English and Japanese sentences (5.4). Section 5.6 describes the methods employed by the investigators to extract predictions of difficulty from the sentence contrasts. On the basis of the predictions alone, the four contrastive analyses are then evaluated for the consistency with which the investigators were able to obtain similar predictions working independently (5.7).

5.1 Selection of the Contrastive Analyses.

Since no two contrastive analyses are fully equivalent, varying both in the model of linguistic description which represents the "descriptive base" and in the contrastive methodology employed to derive predictions (the investigators examined a large number of contrastive analyses - Japanese/English as well as English/other languages - so as to select representative analyses of significantly different types. In essence, there are three broad classifications of descriptive models: the taxonomic (also known as "structural"), the generative, and the stratificational model. Since no contrastive analyses have been developed employing the last of these three, the investigators selected two taxonomic and two generative based analyses. The taxonomic based analyses represent two quite diverse types, immediate constituent analysis and sector analysis. Immediate constituent analysis, best represented in the work of Charles C. Fries (1952) is a form of analysis that depends to a great extent on contextual definition of word classes, and the ways in which word classes can cooccur in strings. Sector analysis defines successive levels of sub-sentence "sectors" in terms of the roles such sectors play in the next highest level. Sector analysis is best represented in the work of Robert L. Allen (1966). Both taxonomic models are oriented to the evident surface structure of the language being described. Generative models, on the other hand, assume that the surface structures are inadequate representations of the real complexity of language behavior, and that "deep" structure representations (which are related to surface structure by complex operations known as "transformations") are more appropriate. While there are now several quite different generative "schools", the contrastive analyses based on the generative model were developed in the early days of generativism, and follow the model proposed by Noam Chomsky (1957). In the course of this project, the investigators do not intend to reflect on the comparative validity of the various models as adequate reflections of language.

The four model selected are hereafter designated Kleinjans, Jackson, Hashimoto, and Stockwell (after their principle authors), and are described below.

Kleinjans:

Kleinjans, Everett, A Descriptive-Comparative Study Predicting Interference for Japanese in Learning English Noun-head Modification Patterns, (Tokyo: Taishukan), 1959.

The descriptive basis of Kleinjans is a structural slot-filler model in the manner of C. C. Fries' (1952) immediate constituent analysis.

Jackson:

Jackson, Kenneth L., English Middle Adverbs and the Japanese Student, (Tokyo: Taishukan), 1970.

The descriptive basis of Jackson is sector analysis (of the tagmemic analysis family) in the manner of Robert L. Allen (1966).

Hashimoto:

Hashimoto, Mitsuo George, From Japanese to English: A Contrastive Analysis Based on a Transformational Model, (unpublished PhD dissertation, Georgetown Univ.), 1967.

The descriptive basis of Hashimoto is transformational, in the manner of N. Chomsky (1957).

Stockwell:

Stockwell, Robert P., Bowen, J. Donald, and Martin, John W., The Grammatical Structures of English and Spanish (Contrastive Structure Series), (Chicago: Univ. of Chicago Press), 1965.

The descriptive basis of Stockwell is also transformational in the manner of Chomsky (1957). Although both Hashimoto and Stockwell are based on the same descriptive model, their contrastive methodologies differ significantly, as described below (section 5.2).

5.2 Variables in the Predictions of Difficulty.

The primary assumption of any contrastive analysis is the set of variables it designates as being the major factors involved in difficulty causation. The secondary assumption is the set of roles that these designated variables play in difficulty. Most (but certainly not all) contrastive analyses define the relevant variables, and then designate their roles in terms of a hierarchy of difficulty, (a system which takes all possible variable interactions and locates them relative to each other as sources of difficulty). The variables and hierarchy of difficulty for each of the four analyses is given below.

Kleinjans

Three variables are significant, defined (Kleinjans, 203) as:

FORM: "the simplest (or minimum) identificational feature of the...structure, which includes the arrangements involved, and the shape of these elements".
DISTRIBUTION: two things - (1) any variation in FORM, (2) any restrictions placed on "the occurrence of the FORM in the larger structural patterns of the language".
MEANING: "any structural meaning inherent in a particular...pattern".

These variables interplay in a four-level hierarchy:

Level A (easiest): F,D,M all marked (+), i.e., the FORMS being contrasted are judged to be similar, as are also their DISTRIBUTIONS and MEANINGS.
Level B: two of F,D,M are marked (+), the other (-).
Level C: two of F,D,M are marked (-), the other (+).
Level D: F,D,M are all marked (-).

Jackson

Three variables are significant, defined (Jackson, 150) as:

FORM: The word order patterns of the item under analysis (further specified in this study to include its level, sector, and filler in terms of sector analysis).

DISTRIBUTION: shiftability of the item (whether or not the relevant form can move to other locations in the sentence).
MEANING: the dependency relationships that the form has with other forms in the sentence.

It should be noted that the Kleinjans and Jackson variables are defined quite differently despite their equivalence of label. The Jackson variables are utilized in an eight-level hierarchy:

Level A (easiest): F,D,M are all marked (+).
Level B: the MEANINGS of the compared forms dissimilar (-),
    B1: F,D marked (+)
    B2: F marked (+), D marked (-)
    B3: F marked (-), D marked (+)
Level C: the MEANINGS of the compared forms similar (+),
    C1: F marked (+), D marked (-)
    C2: F marked (-), D marked (+)
    C3: F,D both marked (-)
Level D: F,D,M all marked (-).

It should be noted that the Kleinjans and Jackson hierarchies are incompatible in the B and C levels, due to a difference in their theoretical interpretations of interference.

Hashimoto

Two main variables are described (Hashimoto, 34-36);

TRANSLATION EQUIVALENCE: mutual or one-directional.
STRUCTURAL EQUIVALENCE: equivalent or non-equivalent in four ways:

1. Changes of word order
2. Structural change (change in dependency)
3. Missing element in target language
4. Extra element in target language

Although Hashimoto refers to the hierarchy of difficulty suggested by Nida (1964), which ranks the structural differences as being the cause of increasing difficulty in the order above, Hashimoto claims to make no effort to create a hierarchy of difficulty in his predictions. For this reason, it was decided that each investigator should establish a hierarchy of difficulty for the Hashimoto model, so that ultimate comparison with the test data would be more or less comparable for all four analytic types. In so far as Hashimoto stated that such a hierarchy was beyond the scope of his work (p. 37), the principles of the hierarchy were left to each investigator independently, within the scope of the Hashimoto variables.

The hierarchies derived are:

(1) With the structural differences as defined above, each kind of difference was associated with a numerical value:
    change of word order: 1 point
    structural change: 2 points
    missing item in TL: 3 points
    extra item in TL: 4 points
In each contrast, the apparent structural differences are noted and their values summed, the total representing the item's level of difficulty.

(2) Numerical values associated with the structural changes:
    change of word order: 1 point
    structural change: 2 points
    missing OR extra item in TL: 3 points
    missing AND extra item in TL: 4 points
In each contrast, the most-highly valued difference is noted and its value represents the item's level of difficulty.

(3) Each type of structural difficulty is given one point, mutual translatability is valued at -1/2 point, the total for any one item doubled to give whole numbers, and revalued upwards so that the lowest value is 0.

(4) Each type of structural difference is given 1 point, the level of difficulty for an item being represented by the sum of the differences.

(5) An approximate valuation of Hashimoto's apparent hierarchy, from his comments in the discussion section following each contrast. 0="no problem", "no difficulty", etc., 1="some difficulty", "slight problem", etc., 2="definite problem", "difficult", etc., and 4="very great problem", "exceptionally difficult", etc.

To illustrate the differences among the above hierarchies, the contrast of the English sentence "The book which I am reading is interesting" and the Japanese sentence "Watakushi ga yonde iru hon wa omoshiroi" was judged by each investigator to involve both a word order change and an extra item in the English sentence. Hierarchy (1) placed it at level 5, (2) at level 3, (3) at level 6, (4) at level 4, and Hashimoto, in his discussion of the relative transformation, says "This is a difficult TR for the Japanese student to learn", i.e., level 2.

Stockwell

Three variables are described (Stockwell, p. 283):

CHOICE: three types:
1. No choice - element lacking in one of the languages
2. Optional choice -
3. Obligatory choice -

STRUCTURAL SIMILARITY
FUNCTIONAL/SEMANTIC SIMILARITY

The concepts OPTIONAL and OBLIGATORY are nowhere defined precisely in Stockwell, nor in the companion volume (Stockwell and Bowen, 1965). In both texts, however, discussion implies that optionality is an absense of restriction by prior context, and obligatoriness is the presence of restriction. STRUCTURAL SIMILARITY is a matter of "same word order, same categories represented" (p.283), and FUNCTIONAL/SEMANTIC SIMILARITY is described in these terms: "though different in word order, the sentences match one-for-one in having corresponding items as subject-verb-object" (p. 283).

The Stockwell variables define a sixteen-level hierarchy, (cf. Table 5.1) with level one representing the level of greatest difficulty.

TABLE 5.1

Levels of Difficulty according to Stockwell

Level	NL CHOICE	STRUC SIMIL	FUNC/ SEMAN	TL CHOICE	
1	Ø	-		OB(ligatory)	(Most difficult)
2	Ø	-		OP(tional)	
3	OP	-	+	OP	
4	OB	-	+	OB	
5	OP	-	+	OB	
6	OB	-	+	OP	
7	OB	-		Ø (No choice)	
8	OP	-		Ø	
9	OP	+	-	OP	
10	OB	+	-	OB	
11	OP	+	-	OB	
12	OB	+	-	OP	
13	OP	+	+	OB	
14	OB	+	+	OP	
15	OP	+	+	OP	
16	OB	+	+	OB	(Easiest)

(After Stockwell, Bowen, Martin, 1965, 284)

The sixteen levels are divided, as indicated, into five major level classes, within which distinctions are made by the interplay of OP and OB.

Since the Stockwell model required that judgements of optionality be made for each sentence, three investigators were assigned the task of so doing independently. Specifically, they were asked to judge whether the filling-in of each blank space on the LSD would involve optionality on the structural level in terms of the transformational model. This task involved determining whether or not all of the possible fillers would have the same transformational description. The results of the exercise (for the detailed report and analysis, see Appendix D) indicated that the investigators' independent motivations for marking optionality were only randomly related to each other. Unable to obtain clear direction from a reading of Stockwell, it was decided that the OP/OB variable could not sensibly be used in this study, and that the Stockwell hierarchy would be collapsed to the five major levels within which OP/OB made finer distinctions. This hierarchy is given in Table 5.2, (reordered so that the highest level is the easiest, to accord with the other hierarchies).

TABLE 5.2

Modified Stockwell Levels of Difficulty

Level	NL	STRUC	FUNC/	TL	
1		+	+		(Easiest)
2		+	-		
3		-		Ø	
4		-	+		
5	Ø	-			(Most difficult)

5.3 Sentence Items.

The contrasts according to the various analyses described above apply to the sentences on the Language Sampling Device, and the prediction is that of the relative difficulty that might be expected of the Japanese students in correctly filling in the LSD blanks.

Since each contrast necessarily involves both an English and a Japanese sentence, it was immediately recognized that there would be a problem in determining an appropriate set of Japanese sentences for the contrasts. The investigators first considered asking informants to provide translations for the LSD items as they were (i.e., with blank spaces, for example: to translate "I do ______ believe it" into Japanese), but it was decided that this task would place too great a demand of the informants. In some cases, the task might not be unreasonable, but for a large number the task of "translating" would almost certainly be meaningless. A second possibility was considered: asking for translations of all the possible sentences that could be reasonably expected of the filled-in LSD sentences. This was rejected on the basis of the fact that (a) it was not known what the full range of possible fillers was and (b) that in any case this would derive such a large number of Japanese sentences as to make analysis far too time consuming a job to complete within the time limits permitted. It was decided, finally, to use only translations of the original sentences from which the LSD was derived.

The implications of this decision cannot be fully assessed. A prediction based on a specific sentence cannot (in theory) be confidently applied to a different sentence, so that, to a certain extent, the predictions that are based on the original filler cannot be said to have general force over all possible fillers. Nevertheless, since in the construction of the LSD (as described in section 3.7 and Appendix C) only items were used that showed a high consistency of categorical sameness in the fillers, it was felt that the same consistency of response might be expected of the Japanese students, at least enough so to minimize the effects of variation. This problem does not effect the LST, of course, since the Japanese students are constrained to one correct answer in that test.

Each sentence was independently translated by three native speakers of Japanese who taught English at the University of Hawaii. These three translations were "normalized" by the investigators - that is, differences in the translations that could not be expected to affect the contrast were merged. Thus, desu, de aru, and da were "normalized" to one form, word-order variations (not involved in the contrast) were eliminated, apparently optional elements were parenthesized, and so on. Where differences were considered potentially significant in terms of one or another of the analyses, they were preserved. In some cases, the result was a single Japanese sentence for contrast, while in other cases, three different sentences had to be contrasted with the English equivalent. The English and Japanese sentences can be found in Appendix E.

5.4 Sentence Analyses.

The English and Japanese sentences were given rough descriptions according to the descriptive base of each analysis. For the Kleinjans, Hashimoto, and Stockwell analyses, each investigator was expected to supplement the rough description with his own knowledge of the linguistic analysis involved and his understanding of the methods and motivations of the person who designed the contrastive method. Jackson provided a detailed sector analysis description of all English and Japanese sentences (since none of the other investigators were familiar with sector analysis), and some instruction in the methods employed in contrasting them. Examples of the descriptions for the sentences can be found in Appendix F.

5.5 Predictions.

Each investigator independently proceeded to extract predictions of difficulty for each sentence for each of the four analyses. These predictions were cast in the form of the hierarchies which the analyses' authors used, or, for the Hashimoto study, in terms of the hierachy designed by the investigator himself (cf. section 5.2).

In cases where there were several Japanese sentences that must be contrasted with a single English contrast, the investigators carried out all contrasts. Where the results were the same, i.e., where the level of difficulty for the sentence was the same for all of the Japanese sentences, that prediction was used. Otherwise, it was left to the discretion of the investigator to decide which level was appropriate or to list the item without any prediction at all.

Since the Kleinjans, Jackson, and Stockwell hierarchies were determined by their authors, a "combined" prediction was obtained for each sentence for these three analyses in the following way: if two or more of the three investigators agreed on the level of difficulty, that level was designated the "average" prediction.

Thus each sentence was associated with three predictions of a level of difficulty, X1, X2, X3 (assigned by the three investigators), and most of the sentences had a fourth, Xa (the "combined"), for each of the Kleinjans, Jackson, and Stockwell systems. The Hashimoto system had five predictions, X1,...,X5, the fifth being Hashimoto's set. These can be found in tabular form in Appendix G. For example, Table 5.3 shows the level of difficulty predictions assigned to sentence 1.

TABLE 5.3

Level of Difficulty Predictions (X) for each Contrastive Analysis, by the Investigators

	Kleinjans		Jackson		Hashimoto		Stockwell	
S#	1 2 3	a	1 2 3	a	1 2 3 4	5(H)	1 2 3	a
1	3 3 3	3	8 8 8	8	3 4 7 3	1	4 4 1	4

In addition, from each investigator's set of predictions of level of difficulty (X), a gross prediction of difficulty (e) was derived. The lowest level of difficulty (X) was considered a prediction of "no difficulty", for which e = 0, and all other levels of difficulty were considered "difficulty", for which e = 1. For example, investigator 1, in the Kleinjans system, assigned each item to one of four levels of difficulty, 1 to 4. Since Level 1 is the lowest level (easiest), all sentences which were assigned X = 1 were then assigned e = 0. Items for which X > 1 were assigned e = 1. All e values are listed in Appendix H.

X and e are the predictive variables that will be correlated to the actual difficulty found in the Japanese performance on the LSD and LST. These correlations and their analysis and interpretation will be found in Chapter 6.

5.6 Measures of Internal Consistency and Objectivity.

One aspect of the evaluation of a contrastive analysis is the ease with which independent investigators can apply it to linguistic data, and the confidence that their results would be consistent with other investigators' results on the same data. The "ideal" contrastive analysis is one that could be machine programmed, and would be wholly objective, requiring no subjective or intuitive response on the part of the investigator. There is, however, considerable reason to believe that a fully objective contrastive analysis has the same status as machine translation - i.e., that for theoretical reasons such objectivity is probably beyond possibility. To a certain extent, contrastive analysis must always be dependent on the talent of the investigator. Obviously, however, it cannot be the case thtt a particular analysis can be highly valued if only its inventor can make it work to any degree of success. In this section, the predictions made by the various independently working investigators are compared to determine the degree to which their predictions are consistent with each other.

For each analysis, rank-order correlation coefficients were obtained for every pairing of investigators, with the Spearman rank-difference formula, and averaged . The coefficients are reported in Table 5.4.

TABLE 5.4

Rank-order Correlation Coefficients between the Investigators' Predictions of Difficulty Levels

Kleinjans

	2	3
1	.76	.57
2	1.00	.59

Ave: .65

Jackson

	2	3
1	.70	40[b]
2	1.00	.65

Ave: .60

Hashimoto

	2	3	4	5(H)
1	.73	.77	.62	.68
2	1.00	.85	.49	.60
3		1.00	.46	.39
4			1.00	.69

Ave: .69 (not incl. H)

Stockwell

	2	3
1	.54	.31[a]
2	1.00	.00

Ave: .30[a]

a - p=.05
b - p=.01 Otherwise, p=.001

As indicated in Table 5.4, the Stockwell analysis has a very low index of prediction consistency, while the other analyses do comparatively well. No data is available, nor is there any theoretical means for appraising the objectivity of the Kleinjans, Jackson, and Hashimoto studies on the basis of the correlation coefficients reported above. In other words, it is not possible to claim that while they do better than the Stockwell they are nevertheless inadequate

---

[1] Since it is not correct procedure to average correlations of coefficients directly, the mean correlation coefficient to Fisher's z function, averaging, and then reconverting. See Quinn McNemar, Psychological Statistics, New York: John Wiley and Sons, 1962, p. 140.

or adequate. It can be seen, however, that the level of consistency with which predictions are made fall well short of the level of the consistency with which difficulty is found - .85 and .89 for the LSD and LST respectively among the eleven groups of Japanese tenth graders (see Chapter 3) - and it may be felt that predictive consistency should at least attain the level of the consistency of the phenomena it is predicting. In this case, of course, none of the analyses may be said to be adequate.

5.7 Summary

Four contrastive analyses were selected for evaluation in this project: Kleinjans, a taxonomic model employing immediate constituent analysis, Jackson, a taxonomic model employing sector analysis, Hashimoto, a generative model whose methodology is strictly related to transformational operations, and Stockwell, a generative model whose contrastive methodology is nevertheless strongly surface-structure oriented.

The variables that are deemed significant by each of the analyses are:

Kleinjans: form, meaning, and distribution, each defined according to an immediate constituent model;

Jackson: form, meaning, and distribution, each defined according to a sector analysis model;

Hashimoto: translatability and structural differences (of which four types are defined transformationally);

Stockwell: choice, structural similarity, and functional/semantic similarity.

Each analysis employs a hierarchy except Hashimoto, for which five hierarchies were independently developed, four by the investigators and the fifth based on Hashimoto's contrastive notes. (These hierarchies are not repeated in this summary, and may be found in section 5.2).

Japanese sentences equivalent to the English sentences on the LSD were obtained from Japanese informants at the University of Hawaii. All English and Japanese sentences were described in terms of each analysis, and the investigators applied the contrastive methodology of each analysis to the sentences to obtain predictions of difficulty levels for each. These predictions will be correlated with the patterns of Japanese errors (cf. Chapter 3) in Chapter 6 to obtain indices for evaluating the predictive capacity of the four analyses.

The four analyses were then evaluated for the consistency by which each analysis ranked the forty items of the LSD in terms of difficulty levels. Correlating the rankings achieved by the different investigators in each analysis, it was found that the Stockwell analysis had an average correlation coefficient that was only barely significant (.30), and that the other three achieved levels between .60 and .69, much better than the Stockwell, but not necessarily particularly good. It was concluded that the objectivity of all the analyses is deficient, the Stockwell especially so.

CHAPTER VI. THE POWER OF CONTRASTIVE ANALYSES TO PREDICT DIFFICULTY

6.0 In this chapter, the power of contrastive analyses to predict difficulty will be evaluated. Four indices of predictive capacity will be derived by correlating the predicted levels of difficulty with the actual performance of the Japanese students on the LSD and the LST. As was noted in Chapter 3, analysis of the perfomance similarities among the various Japanese students indicated that between 40% and 60% of the performance similarities may be accounted for by the two commonalities of their all being native speakers of Japanese, and their all learning English under the Japanese educational system. It was also noted that it was not possible to distinguish between these two factors. Nevertheless, it may be presumed that the native language commonality accounts for some portion of the 40% - 60%, and it is that portion that the contrastive analyses being evaluated are presumed to be predicting. In other words, there is apparently room within which the contrastive analyses can perform significantly if they are in fact adequate.

The predictive variables that go into the indices are X and e, the predictions of level of difficulty and gross predictions of difficulty respectively (see Chapter 5 for full descriptions of each, and Appendices G and H for tables of X and e).

The performance variables that go into the indices are E (and its variants E_1, E_2, E_3), P, and T, or gross performance difficulty, percentage correct, and tenacity of difficulty respectively (see Chapter 3 for full description of each).

6.1 The Indices.

Four indices of predictive power are calculated in the form of Pearson product-moment correlation coefficients between the rank-orders predicted by e and X and the actual rank-orders as indicated by E (and its variants), P, and T. In the tables below, no correlation smaller than ±0.257, the minimum correlation that is significant at the 0.05 level for a one-tailed test with forty degrees of freedom, will be reported. In such a case, three dots (...) will be employed. Otherwise, the reportage in the tables will reflect the following three possible events.

1. an underlined coefficient will be used to indicate the three investigators' correlations averaged together, if this is above ±.257.

2. a simply-reported coefficient will indicate that the first case does not apply, but that the "combined" prediction (cf. 5.5) achieved significant value.

3. a coefficient reported in parentheses indicates that neither of the above cases apply, but that one of the investigators' predictions correlated significantly.

The above three possible events are understood to be in decreasing order of value. In the first case, all three investigators probably score in the "significant" range. In the second, the "pooling" of predictions is seen as having a resultant significance, whereas in the third event, a single investigator stands isolated. As it turns out, the three events are generally implicatorily related, i.e., if event 1 is applicable, events 2 and 3 were also, and so on. As will be seen below, the analyses will be evaluated in terms of their performance in achieving significant results in the above three cases.

Gross capacity of Prediction (eE, eE_1, eE_2, eE_3). These correlation coefficients represent various calculations of the relationship between the gross predictions of difficulty and the gross occurrence of difficulty, and are found in Table 8.1.

TABLE 6.1

Gross Capacity of Prediction

	LSD				LST			
Analysis	eE	eE_1	eE_2	eE_3	eE	eE_1	eE_2	eE_3
Kleinjans	...	...	...	...	...	...	...	...
Jackson	...	...	...	...	...	...	-.332	-.376
Hashimoto	...	...	...	...	...	...	...	...
Stockwell	...	...	(.331)	.333	...	...	...	...

The main conclusion concerning the gross capacity of contrastive analyses to predict difficulty is that it hardly exists. Those few significant correlations are only barely significant, and there are no cases of the most highly valued event (in which all investigators score significantly). The Jackson and the Stockwell systems do approximately equally well, but it is important to note that the Jackson coefficients are negative while the Stockwell coefficients are positive. To the degree that the results are significant, they are incompatible as well. The negative results with the *Jackson* system imply that the hierarchy of difficulty must be wholly reversed - i.e., that similarity, and not difference, is the major element causing difficulty.

Relative gross capacity of Prediction (eP). Since the relative difficulty is a more finely tuned version of *gross capacity*, the values are expected to be lower generally. It was not surprising, therefore, to discover that in the entire set of eP correlations, only one investigator achieved a significant score, as reported in Table 6.2. Under the circumstances, with respect to this measurement no analysis may be said to have achieved any significant capacity. (That the one correlation is negative is consonant with expectation, since P is actually ordered opposite to e).

TABLE 6.2

Relative Gross Capacity (eP) of Prediction

Analysis	LSD	LST
Kleinjans	...	...
Jackson	...	...
Hashimoto	...	...
Stockwell	(-.295)	...

Capacity to Predict Relative Difficulty (XP). This index represents the relative predictive capacity of a contrastive analysis or, in particular, the value of its hierarchy of difficulty as representing real levels of difficulty. Since P is actually the percentage of correct response, such that the higher the value the easier the item is presumed to be, and since X is ordered oppositely - successive levels indicate increasing difficulty, the expected correlation coefficients are negative. The results are reported in Table 6.3.

TABLE 6.3

Capacity to Predict Relative Difficulty (XP)

Analysis	LSD	LST
Kleinjans	...	(-.273)
Jackson	.266	.342
Hashimoto	...	...
Stockwell	(-.287)	...

Only the Jackson system may be said to achieve significant results, albeit only slightly greater than the minimum level reported. Again, the Jackson coefficients are opposite in sign to those expected.

Capacity to predict Tenacity (XT). While no analysis claims to measure the possible "tenacity" of the difficulty of an item, it was felt that since there is a demonstrable relationship between P and T (with correlations between P and T of -.476 on the LSD and -.560 on the LST), a predictive hierarchy might be expected to correlate with tenacity. At any rate, since the hierarchies relate so poorly to real difficulty, they might do better with tenacity. The results are reported in Table 6.4.

TABLE 6.4

Capacity to predict Tenacity (XT)

Analysis	LSD	LST
Kleinjans	...	...
Jackson	-.449	(-.271)
Hashimoto	...	...
Stockwell	...	(-.298)

Once again, the results are generally poor and only the Jackson system can be given any credit for attaining significant predictive capacity. Once again, it must be noted that the coefficients are opposite in sign to those expected.

As a summary of the evaluation of the predictive performance of the various analyses, point values were associated with the three event types: three points for event 1, two points for event 2, and one point for event 3 (see section 6.1 for a definition and explanation of these events). The points achieved by any one analysis in all the tests were summed (with a possible maximum of 42 points), and are reported in Table 6.5.

TABLE 6.5

Contrastive Analysis Performance Points

Analysis	Points
Kleinjans	1
Jackson	12
Hashimoto	0
Stockwell	6

Honors must be given to the Jackson system of contrastive analysis, second place to Stockwell, and Kleinjans and Hashimoto trail badly.

6.2 Conclusions.

No analysis demonstrates a consistent ability to predict linguistic performance to a high degree. There are indications that, however, one of the basic assumptions - that difficulty and linguistic difference are directly related - is incorrect.

What is perhaps most notable about the Jackson results is the fact that every correlation coefficient is opposite in sign to those expected. This fact, coupled with the fact that the Jackson system scored best is suggestive of the possibility that one of the basic tenets of contrastive analysis is wrong, that "difference" does not, in fact, correlate with "difficulty". If the Jackson results are significant, they can only be interpreted in the light that, somehow, relative similarity correlated with relative difficulty. This does not, of course, account for the fact that the other analyses did poorly relative to Jackson. One possible explanation for this is that the Jackson hierarchical variables were better motivated than the other analyses' variables (i.e., were more closely matched to the presumptive 'real' variables involved in linguistic transfer). Another possible explanation is the greater value of sector analysis in contrastive analysis. No attempt is made in this project to analyze the differences among the four analyses, however, since even if Jackson is "better", it is not really very good, having a sporadic and generally low-level performance.

6.3 Summary.

This chapter concerned the relationships between the predictions obtained from the four contrastive analyses and the data on performance difficulty among the Japanese subjects. Four indices were calculated, all on the basis of Pearson product-moment correlation coefficients. These indices are:

1. (eE) Index of gross capacity to predict gross difficulty,
2. (eP) Index of gross capacity to predict relative difficulty,
3. (XP) Index of capacity to predict relative difficulty,
4. (XT) Index of capacity to predict error tenacity.

Correlation coefficients were reported only if they were greater than ±.257, below which they are nonsignificant for a two-tailed test at p = .05, and were reported in three grades, the most valuable of which was the case in which the three investigators' correlation coefficients averaged more than ±.257, the next in which the "combined" coefficient was greater than ±.257, and the last of which being the case in which only one of the three investigators achieved a coefficient above ±.257.

In the indices so calculated, there was only one occurrence of the most highly valued case, five in which the "combined" coefficient was significant, and six individual investigators' significant coefficients. Of these, only one was higher than .400, five fell between .300 and .400, and six between .257 and .300. In general, the analyses' performances were sporadic and only barely above the minimum level of significance.

Of the four analyses, only the Jackson system could be said to be consistent, having the lion's share of reported correlations (fully half). It may be significant that all of the correlation coefficients reported under the Jackson system are opposite in sign to those expected. The only way to interpret this fact is to hypothesize that the Jackson hierarchy must be revalued, such that the direction of increasing difficulty is in the opposite direction to the one hypothesized by Jackson. This may mean that difficulty is not, in fact, directly related to linguistic differences, but instead to linguistic similarity.

If this were the case, however, then one would expect the other analyses to reflect the same reversal in sign, which is not consistently so. There is, simply, insufficient data by which to speculate further.

In sum, even the Jackson system must be considered as being sporadic and performing at a relatively low level. The main conclusion of this chapter is that contrastive analyses are still at so primitive a stage that they are unable to perform well predictively.

CHAPTER VII. CONCLUSIONS AND RECOMMENDATIONS

7.0 This study attempted to test three aspects of the problem of validity of contrastive analysis as a means for predicting errors or problems for second language learners: the constancy of foreign-language errors, the objectivity of the methods and procedures of contrastive analysis, and the capacity of contrastive analyses to make accurate predictions. The conclusions from this study related to these three aspects of the problem of validity reported in turn below are both discouraging to practicioners of contrastive analysis and encouraging to contrastive theorists.

7.1 Errors were found to be fairly constant for Japanese learners of English in spite of differences in geographic, socio-economic, dialectal, pedagogical, etc. backgrounds. This suggests that the output of a contrastive analysis can be generalized for all speakers of a given language even if only a single dialect of the source language is contrasted with a single dialect of the target language.

Errors were also found to decrease in number through time. The intrinsic English difficulty of the tests plays a significant role, accounting for between 15% and 35% of the commonality of difficulty among the subjects, after two or three years of English study --the point at which the performance of the Japanese subjects matched that of the American subjects. At this point, common language and culture and a common educational system appear to be more important factors in error production, accounting for 40% to 60% of the commonality of performance.

The factor of 'ignorance' of English which is assumed to be the major aspect of differing levels, plays what may be a dominant role in the first two years of English study, but seems to disappear as a significant factor in later acquisition. This suggestion is based on so little data, however, as to be almost wholly unsubstantiated, and requires a great deal more research before it can confidently be asserted. Even so, items which are difficult initially tend to remain so for a period of six years. This suggests that the predictions made by a contrastive analysis would be as adequate for the first year learner as for the sixth.

7.2 The methods and procedures of contrastive analysis were found to be deficient in objectivity in the sense that the methods could be replicated by other investigators. While total objectivity is, perhaps, like machine translation, an impossibility at the present time, this finding suggests that extant contrastive analyses are fairly subjective and are only able to make correct generalizations about the errors of language learners if the analyst has some *a priori* knowledge of what those errors are. In any case, the output of a contrastive analysis remains a hypothesis about errors the student may make and must be compared with his actual usage.

7.3 In terms of the capacity of contrastive analyses to make accurate predictions, it must be concluded that the contrastive analyses examined failed utterly to predict the problems that Japanese students would have on either of the tests administered, at least in terms of the analyses' given assumptions.

The Jackson method, however, achieved a certain adequacy --albeit sporadic and minor--but only if the basic contrastive assumptions that "difference causes difficulty" is reversed. Since the Jackson correlations were minor (seldom being greater than 0.35) and sporadic, it is impossible to conclude firmly that, in fact, this theoretic reversal is valid. Nevertheless, there is no substantiation whatsoever for the concept that linguistic difference is a cause of difficulty.

If the reversal is a valid one, in any case, there is no direct explanation for the fact that only the Jackson method achieved significant predictive adequacy and none of the others did. It may be that the explanation lies in the fact that the Jackson variables were nearly untranslatable into the terms of the variables of the other analyses, and that, say, one or another of the Jackson variables somehow reflected a real factor involved in foreign language difficulty. If this is the case, the data do not provide exploration into the nature of the variables that might have so reflected some aspect of language learning reality. This is without question, an area that demands additional research.

In effect, this study seems to indicate that contrastive analysis can make, in theory, a large contribution to the preparation of materials for second-language learning, but that the state of the art today is so primitive that, in practice, it does not. What is required is substantial research into the very basics of the theory of the mechanics of language interference, and into the relevant variables that must play a role in interference.

7.4 On the basis of the findings of this study, the investigators make the following recommendations:

1. Materials for second-language learning should be based on descriptions of the target language and not on the output of contrastive analyses (until such time as contrastive linguistic theory becomes tenable);

2. More testing of "vertical" grades in Japan and testing in a number of areas where native speakers of other languages may be cross-correlated with the Japanese and American should be carried out, perhaps, to uncover some universal bases for error production.

3. Factor analyses of the variables accounting for commonality of errors for learners with the same amount of learning experience should be carried out in order to tease out the degree to which linguistic rather than psychological or sociological factors affect second-language learning, if at all.

4. Factor analysis of the variables related to ignorance of the target language is indicated in order to assess the point in second language acquisition when native language competence begins to interfere.

5. Extensive basic research is needed in the mechanics of language interference and the variables that play a role in it.

REFERENCES

1. Allen, Robert L., THE VERB SYSTEM OF PRESENT-DAY AMERICAN ENGLISH, The Hague: Mouton & Co., 1966.

2. Bendix, Edward Herman, "Componential Analysis of General Vocabulary," IJAL 32, No. 2, 1966.

3. Bever, Thomas G., "The Cognitive Basis for Linguistic Structures," in John B. Hays, ed., COGNITION AND THE DEVELOPMENT OF LANGUAGE, New York: John Wiley and Sons, 1970, pp. 279-362.

4. Carroll, John B., "Contrastive Linguistics and Interference Theory," in James E. Alatis, ed., REPORT OF THE NINETEENTH ANNUAL ROUND TABLE MEETING ON LINGUISTICS AND LANGUAGE STUDIES, Washington, D.C.: Georgetown University Press, 1968, pp. 113-122.

5. Carroll, John B., Aaron S. Carton, and Claudia P. Wilds, AN INVESTIGATION OF "CLOZE" ITEMS IN THE MEASUREMENT OF ACHIEVEMENT IN FOREIGN LANGUAGES, Cambridge, Mass.: Laboratory for Research In Instruction, Graduate School of Education, Harvard, University, 1959.

6. Chomsky, Noam, LANGUAGE AND MIND, New York: Harcourt, Brace and Company, 1968.

7. Chomsky, Noam, SYNTACTIC STRUCTURES, The Hague: Mouton & Co., 1957.

8. Crow, Edwin L., Francis A. Davis, and Margaret W. Maxfield, STATISTICS MANUAL, New York: Dover Publications, 1960.

9. Dingwall, William Orr., "Transformational Generative Grammar and Contrastive Analysis," LANGUAGE LEARNING, 14:147-160, 1964.

10. Ferguson, Charles A., "General Introduction to the Series," in William Moulton, THE SOUNDS OF ENGLISH AND GERMAN, Chicago: University of Chicago Press, 1962, pp. v-vi.

11. Fries, Charles C., THE STRUCTURE OF ENGLISH, New York: Harcourt, Brace and Company, 1952.

12. Gradman, Harry, "The Limitations of Contrastive Analysis Predictions," WORKING PAPERS IN LINGUISTICS, Vol. 3, No. 4, April 1971, pp. 73-77.

13. Harris, David P., TESTING ENGLISH AS A SECOND LANGUAGE, New York: McGraw-Hill Book Company, 1969.

14. Hashimoto, Mitsuo, "From Japanese to English," Unpublished Ph.D. dissertation, Georgetown University, Washington, D.C., 1966.

15. Jackson, Kenneth L., ENGLISH MIDDLE ADVERBS AND THE JAPANESE STUDENT, Tokyo: Taishukan, 1970.

16. Jakobovits, Leon A., FOREIGN LANGUAGE LEARNING, Rowley, Mass.: Newbury House Publishers, 1970.

17. Kleinjans, Everett, A DESCRIPTIVE-COMPARATIVE STUDY PREDICTING INTERFERENCE FOR JAPANESE LEARNING ENGLISH NOUN-HEAD MODIFICATION PATTERNS, Tokyo: Taishukan, 1959.

18. Kokugo Gakkai, KOKUGO GAKKAI KOKUGOGAKU JITEN, Ed. Tokieda Motoki, Tokyo: Tokyodo, 1955.

19. Kucera, Henry and William Francis, COMPUTATIONAL ANALYSIS OF PRESENT-DAY AMERICAN ENGLISH, Providence, R.I.: Brown University Press, 1967.

20. Lado, Robert, LANGUAGE TEACHING: A SCIENTIFIC APPROACH, New York: McGraw-Hill, Inc., 1964.

21. __________, LANGUAGE TESTING, New York: McGraw-Hill Book Co., Inc., 1961.

22. __________, LINGUISTICS ACROSS CULTURES, Ann Arbor: University of Michigan Press, 1957.

23. Lado, Robert, Charles C. Fries, et al., AN INTENSIVE COURSE IN ENGLISH, 4 vols., Ann Arbor: University of Michigan Press, 1958.

24. Lee, W. R., "Thoughts on Contrastive Linguistics in the Context of Language Teaching," in James E. Alatis, ed., REPORT OF THE NINETEENTH ANNUAL ROUND TABLE MEETING ON LINGUISTICS AND LANGUAGE STUDIES, Washington, D.C.: Georgetown University Press, 1968, pp. 185-194.

25. MacIntyre, Alasdair, "Noam Chomsky's View of Language," in Mark Lester, ed., READINGS IN APPLIED TRANSFORMATIONAL GRAMMAR, New York: Holt, Rinehart, and Winston, Inc., 1970, pp. 96-113.

26. McNemar, Quinn, PSYCHOLOGICAL STATISTICS, New York: John Wiley and Sons, Inc., 1962.

27. Nida, Eugene, TOWARD A SCIENCE OF TRANSLATING: WITH SPECIAL REFERENCE TO PRINCIPLES AND PROCEDURES, Leiden: E. J. Brill, 1964.

28. Osgood, C. E., "The Similarity Paradox in Human Learning: A Resolution," PSYCHOLOGICAL REVIEW, LVI., pp. 132-154, 1949.

29. Politzer, Robert L., TEACHING FRENCH: AN INTRODUCTION TO APPLIED LINGUISTICS, Boston: Ginn and Co., 1960.

30. Politzer, Robert L., Charles N. Staubach, TEACHING SPANISH: A LINGUISTIC ORIENTATION, Waltham, Mass: Blaisdell and Co., 1965.

31. Ritchie, William C., "Some Implications of Generative Grammar for the Construction of Courses in English as a Foreign Language," LANGUAGE LEARNING, 17: 45-131, 1967.

32. Rivers, Wilga M., THE PSYCHOLOGIST AND THE FOREIGN LANGUAGE TEACHER, Chicago: University of Chicago Press, 1964.

33. Roberts, Paul, CORSO D'INGLESE PARLATO, 3 vols., New York: Harcourt, Brace and World, 1963-1965.

34. Siegel, Sydney, NON-PARAMETRIC STATISTICS FOR THE BEHAVIORAL SCIENCES, New York: McGraw-Hill, 1956.

35. Stern, Jess, ed., THE RANDOM HOUSE DICTIONARY OF THE ENGLISH LANGUAGE, New York: Random House, 1966.

36. Stockwell, Robert, "Contrastive Analysis and Lapsed Time," in James E. Alatis, ed., REPORT OF THE NINETEENTH ANNUAL ROUND TABLE MEETING ON LINGUISTICS AND LANGUAGE STUDIES, Washington, D.C.: Georgetown University Press, 1968, pp. 11-26.

37. Stockwell, Robert, David J. Bowen, and John W. Martin, THE GRAMMATICAL STRUCTURES OF ENGLISH AND SPANISH, Chicago: University of Chicago Press, 1965.

38. Stockwell, Robert, Donald J. Bowen, THE SOUNDS OF ENGLISH AND SPANISH, Chicago: University of Chicago Press, 1960.

39. Wardhaugh, Ronald, "The Contrastive Analysis Hypothesis," TESOL QUARTERLY, 4:123-130, 1970.

40. ________, "Some Current Problems in Second-Language Teaching," LANGUAGE LEARNING, 17:21-26, 1967.

41. Thorndike, E. L., ed., THORNDIKE CENTURY JUNIOR DICTIONARY, Rev. ed., Chicago: Scott, Foresman and Co., 1942.

42. West, Michael, ed., A GENERAL SERVICE LIST OF ENGLISH WORDS WITH SEMANTIC FREQUENCIES AND A SUPPLEMENTARY WORD-LIST FOR THE WRITING OF POPULAR SCIENCE AND TECHNOLOGY, Rev. ed., London: Longmans, 1953.

LANGUAGE SAMPLING TEST

The examiner will read these instructions. Please do not open your test booklet until he has finished.

INSTRUCTIONS:

1. Fill in the information asked for on the separate answer sheet.
2. Then, read the following instructions carefully, and be sure you understand them fully before you begin the test.
3. If you have any questions, ask the examiner before the test begins.
4. There are forty sentences with blanks in them in this Language Sampling Test. Under each sentence, there are four words and you are to choose the word which will make the sentence a good English sentence.
5. There is only one correct answer for each sentence. Blacken only one space on the answer sheet with the special pencil the examiner will give you. If you make a mistake or want to change an answer, be sure to erase the mistake very carefully.
6. Please do not write in this test booklet. Mark your answers in the proper column on the answer sheet.
7. Please answer all of the problems.
8. You don't have much time so do not spend a lot of time on one problem or you will not have time to complete all forty items.

EXAMPLE:

E1 The boy ____________________ the house.
(a) sees
(b) see
(c) seen
(d) is

The correct answer for E1 is "(a) sees." To show that "(a)" is the correct answer, blacken the space under "(a)" on your answer sheet.

Check to see that your name is on your answer sheet. Then, open your test booklet and begin the test.

Page One

1. I can ____ see him.
(a) no
(b) not
(c) ever
(d) yet

2. Mary ____ to school with me last week.
(a) comes
(b) went
(c) was
(d) go

3. They don't have ____.
(a) any
(b) none
(c) no
(d) yet

4. Her ____ is on the table.
(a) always
(b) hands
(c) pretty
(d) book

5. John ____ know her.
(a) also
(b) no
(c) doesn't
(d) not

6. She ____ waiting for us.
(a) is
(b) may
(c) be
(d) are

7. We saw ____ last week.
(a) us
(b) they
(c) him
(d) your

8. Are you ____ to New York?
(a) go
(b) went
(c) gone
(d) going

9. I ____ that movie last night.
(a) seen
(b) went
(c) saw
(d) gone

10. I read a ____ book.
(a) teacher
(b) novel
(c) child
(d) new

11. They ____ to visit us in France.
(a) might
(b) used
(c) has
(d) suppose

12. ____ she feel well?
(a) Girl
(b) Is
(c) Does
(d) Wills

13. Let's ____ home as soon as possible.
(a) go
(b) went
(c) going
(d) to go

14. ____ he go to school this morning?
(a) How
(b) Is
(c) Had
(d) Did

15. He ____ look at this thing.
(a) should
(b) has
(c) not
(d) do

16. He has lost both ____.
(a) slipper
(b) shoes
(c) two
(d) them

17. ___ he come with me?
(a) Can
(b) Do
(c) Is
(d) No

18. I ___ home at 8:00 last night.
(a) my
(b) go
(c) gone
(d) went

19. The boy ___ seen by his father.
(a) warmly
(b) that
(c) was
(d) will

20. I went to see a ___ movie.
(a) long
(b) crying
(c) today
(d) actor

21. Please talk to ___ about it.
(a) they
(b) him
(c) girl
(d) quietly

22. I don't have much ___.
(a) foods
(b) pencil
(c) water
(d) ideas

23. ___ will they arrive?
(a) What
(b) Time
(c) There
(d) When

24. The school has lost one of ___ teachers.
(a) more
(b) its
(c) school
(d) all

25. I ___ not seen him today.
(a) Have
(b) am
(c) will
(d) do

26. ___ of the boys have come.
(a) No
(b) First
(c) All
(d) Much

27. The girl ___ spoke is my best friend.
(a) having
(b) whom
(c) he
(d) that

28. He ___ John to go to New York.
(a) would
(b) wish
(c) wants
(d) will

29. Read the book ___ you have.
(a) if
(b) which
(c) carefully
(d) who

30. We ___ John walking to his house.
(a) hoped
(b) saw
(c) reminded
(d) would

31. They looked at the ___ boy.
(a) very
(b) run
(c) big
(d) two

32. ___ you like that man?
(a) Do
(b) Have
(c) How
(d) Who

33. She is the girl ___ you saw in school.
    (a) first
    (b) pretty
    (c) friends
    (d) that

34. ___ will you do it?
    (a) What
    (b) Not
    (c) Do
    (d) How

35. My mother is ___ here.
    (a) not
    (b) yet
    (c) no
    (d) she

36. The water ___ finished before we arrived.
    (a) had
    (b) was
    (c) did
    (d) would

37. ___ old are you?
    (a) How
    (b) What
    (c) Twelve
    (d) You

38. John did ___ agree with me.
    (a) must
    (b) no
    (c) he
    (d) not

39. We had ___ that he would come.
    (a) him
    (b) known
    (c) saw
    (d) to

40. Is there ___ order for me?
    (a) new
    (b) of
    (c) an
    (d) first

英語使用試験

[注意] 試験監督が注意事項を読み上げます。それが終るまでこのテストをあけないで下さい。

注意事項

1. 解答用紙にある空白にあなたの名前、日付などを書き入れて下さい。
2. それから次の注意事項をよく読んで下さい。
3. もし質問があればテストが始まる前に試験監督に聞いて下さい。
4. このテストに空白のある英語の文があります。各文の下に単語が四つありますが、その単語の中から正しい単語を一つ選べば文が完全になるようになっています。
5. 正しい答は各文に一つしかありません。解答用紙には各問題について一ケ所だけ監督の渡した特別の鉛筆で黒くして下さい。もし間違つて黒くしたら、消ゴムで完全に消して下さい。
6. 問題用紙には何も書き込まないで下さい。解答は解答用紙の各解答欄につけて下さい。
7. 全部の問題に答えて下さい。
8. テストの時間は四十五分です。一つの問題にあまり時間をかけますと全部解答することができなくなります。

英語使用試験

練習問題

E.1 The boy__________the house.

(a) sees

(b) see

(c) seen

(d) and

E.1の問には(a) sees が正しい答です。(a)が正しいことを表わすために解答用紙にある(a)の欄の ॥ を完全に黒くして下さい。

では名前を記入したかどうかをもう一度確かめてからテストを始めて下さい。

1. I can ___ see him.
   (a) no
   (b) not
   (c) ever
   (d) yet

2. Mary ___ to school with me last week.
   (a) comes
   (b) went
   (c) was
   (d) go

3. They don't have ___.
   (a) any
   (b) none
   (c) no
   (d) yet

4. Her ___ is on the table.
   (a) always
   (b) hands
   (c) pretty
   (d) book

5. John ___ know her.
   (a) also
   (b) no
   (c) doesn't
   (d) not

6. She ___ waiting for us.
   (a) is
   (b) may
   (c) be
   (d) are

7. We saw ___ last week.
   (a) us
   (b) they
   (c) him
   (d) your

8. Are you ___ to New York?
   (a) go
   (b) went
   (c) gone
   (d) going

9. I ___ that movie last night.
   (a) seen
   (b) went
   (c) saw
   (d) gone

10. I read a ___ book.
    (a) teacher
    (b) novel
    (c) child
    (d) new

11. They ___ to visit us in France.
    (a) might
    (b) used
    (c) has
    (d) suppose

12. ___ she feel well?
    (a) Girl
    (b) Is
    (c) Does
    (d) Wills

13. Let's ___ home as soon as possible.
    (a) go
    (b) went
    (c) going
    (d) to go

14. ___ he go to school this morning?
    (a) How
    (b) Is
    (c) Had
    (d) Did

15. He ___ look at this thing.
    (a) should
    (b) has
    (c) not
    (d) do

16. He has lost both ___.
    (a) slipper
    (b) shoes
    (c) two
    (d) them

17. ___ he come with me?
(a) Can
(b) Do
(c) Is
(d) No

18. I ___ home at 8:00 last night.
(a) my
(b) go
(c) gone
(d) went

19. The boy ___ seen by his father.
(a) warmly
(b) that
(c) was
(d) will

20. I went to see a ___ movie.
(a) long
(b) crying
(c) today
(d) actor

21. Please talk to ___ about it.
(a) they
(b) him
(c) girl
(d) quietly

22. I don't have much ___.
(a) foods
(b) pencil
(c) water
(d) ideas

23. ___ will they arrive?
(a) What
(b) Time
(c) There
(d) When

24. The school has lost one of ___ teachers.
(a) more
(b) its
(c) school
(d) all

25. I ___ not seen him today.
(a) have
(b) am
(c) will
(d) do

26. ___ of the boys have come.
(a) No
(b) First
(c) All
(d) Much

27. The girl ___ spoke is my best friend.
(a) having
(b) whom
(c) he
(d) that

28. He ___ John to go to New York.
(a) would
(b) wish
(c) wants
(d) will

29. Read the book ___ you have.
(a) if
(b) which
(c) carefully
(d) who

30. We ___ John walking to his house.
(a) hoped
(b) saw
(c) reminded
(d) would

31. They looked at the ___ boy.
(a) very
(b) run
(c) big
(d) two

32. ___ you like that man?
(a) Do
(b) Have
(c) How
(d) Who

33. She is the girl ___ you saw in school.
(a) first
(b) pretty
(c) friends
(d) that

34. ___ will you do it?
(a) What
(b) Not
(c) Do
(d) How

35. My mother is ___ here.
(a) not
(b) yet
(c) no
(d) she

36. The water ___ finished before we arrived.
(a) had
(b) was
(c) did
(d) would

37. ___ old are you?
(a) How
(b) What
(c) Twelve
(d) You

38. John did ___ agree with me.
(a) must
(b) no
(c) he
(d) not

39. We had ___ that he would come.
(a) him
(b) known
(c) saw
(d) to

40. Is there ___ order for me?
(a) new
(b) of
(c) an
(d) first

APPENDIX B

[illegible] 273

LANGUAGE SAMPLING DEVICE

NAME ______________________________ DATE ________________

SCHOOL ____________________________ AGE ________________

The examiner will read these instructions. Please do not open your test booklet until he has finished.

INSTRUCTIONS:

1. Fill in the information asked for above.
2. Then, read the following instructions carefully, and be sure you understand them fully before you begin the test.
3. If you have any questions, ask the examiner before the test begins.
4. There are forty sentences in the Language Sampling Device. Each sentence has a blank in it. This blank represents a single word which was taken out of the sentence.
5. Try to guess what word was taken out and then write the word you decide on in the blank. BE SURE YOU USE ONLY ONE WORD IN A BLANK. DO NOT USE ANY NAMES.
6. Check your sentence to make sure it is good English and then go on to the next sentence.
7. If you think of several words which could be put in the blank, choose one of them. Maybe you'll guess the word that was taken out of the sentence.
8. Please fill in all of the blanks.
9. Do not spend too much time answering one sentence or you will not have time to complete all forty sentences.
10. The same word may be used in more than one sentence.
11. Remember, USE ONLY ONE WORD IN A BLANK and DO NOT USE ANY PROPER NAMES.

EXAMPLE: The boy ________________________ the house.

1. Many words such as "sees, saw, entered, likes," and so on can be used in this sentence.
2. Words like "lived in" cannot be used because there are two words, "lived" and "in". However, contractions like "don't, can't," and so on may be used.
3. Also, "see" cannot be used in the Example because "The boy see the house." is not a good English sentence.

Be sure your name in on your paper and then please begin the test.

1. I do ______________________________ believe it.

2. I ______________________________ late to school last week.

3. Have you any money? We haven't ______________________________.

4. Her ______________________________ is at 25 South Street.

5. He ______________________________ speak English yet.

6. They ______________________________ waiting.

7. Give ______________________________ a drink.

8. He is ______________________________ a book.

9. She ______________________________ a party last night.

10. John is an ______________________________ person.

11. He ______________________________ to come at 10:00.

12. ______________________________ you help us?

13. Let's ______________________________ a visit to New York.

14. ______________________ she have any?

15. You ______________________ read this story.

16. I worked there three full ______________________.

17. ______________________ I go now?

18. School ______________________ at 9:00 last year.

19. A child ______________________ loved by its mother.

20. I've written a ______________________ story about it.

21. She isn't here. Have you seen ______________________?

22. A book contains much ______________________.

23. ______________________ does school close?

24. We gave one book to ______________________ boy.

25. John ______________________ been sick.

26. He gave me ______________________ of them.

27. Is he the man ______________________ sells food?

28. I ______________________ someone to read it to me.

29. The book ______________________ I am reading is very interesting.

30. I ______________________ him coming.

31. I wrote him a ______________________ letter.

32. ______________________ you come all the way by foot?

33. There are times ______________________ one must be early.

34. ______________________ did it happen?

35. Six and two do ______________________ make ten.

36. The food ______________________ eaten.

37. ______________________ is your name?

38. John thought so, but I ______________________ agree.

39. He had ______________________ an important man.

40. ______________________ man I saw had no feet.

英語使用調査

あなたの名前 ______________________ 年令 ________ 日付 ________

先生の名前 ______________________ 学校名 ______________________

注意 試験監督が注意事項を読み上げます。それが終るまでこのテストをあけないで下さい。

注意事項

1. 上の空白にあなたの名前、年令などを書き入れて下さい。
2. それから次の注意事項をよく読んで下さい。
3. もし質問があればテストが始まる前に試験監督に聞いて下さい。
4. このテストに英語の文が40あります。各文には空白があり、その空白に単語を一つ入れれば文が完全になるようになっています。
5. この空白にあてはまる単語をよく考えて、書き入れて下さい。一つの空白に単語を一つ以上入れてはいけません。人名、地名のような固有名詞は使わないようにして下さい。
6. 単語を書き入れた文が正しい英語かどうかを確かめて下さい。
7. 空白に入れる単語が一つ以上考えられる場合は、その中から適当なものを一つだけえらんで書き入れて下さい。
8. 空白は全部書き入れて下さい。
9. 一つの文にあまり時間をかけないようにして下さい。そうしないと40の文を全部答える時間が足りなくなります。
10. 同じ単語を一度以上使ってもかまいません。
11. もう一度注意しますが、各空白に単語を一つだけ書き入れ、固有名詞は使わないようにして下さい。

練　習　問　題

The boy ______________________ the house.

1. この文の空白には sees, saw, entered, likes のような単語があてはまりますが、その中のどれか一つをえらんで書き入れます。

2. lived in のような句は単語が二つあるので使ってはいけません。しかし、don't や can't のような短縮形は使ってかまいません。

3. 上の例文の空白には see や seen のような単語を書き入れてはいけません。文法的に正しい英語の文にならないからです。

では、名前を記入したかどうかをもう一度確かめてからテストを始めて下さい。

1. I do ____________________ believe it.

2. I ____________________ late to school last week.

3. Have you any money? We haven't ____________________.

4. Her ____________________ is at 25 South Street.

5. He ____________________ speak English yet.

6. They ____________________ waiting.

7. Give ____________________ a drink.

8. He is ____________________ a book.

9. She ____________________ a party last night.

10. John is an ____________________ person.

11. He ____________________ to come at 10:00.

12. ____________________ you help us?

13. Let's ____________________ a visit to New York.

14. ____________________ she have any?

15. You ____________________ read this story.

16. I worked there three full ____________________.

17. ____________________ I go now?

18. School ____________________ at 9:00 last year.

19. A child ____________________ loved by its mother.

20. I've written a ____________________ story about it.

21. She isn't here. Have you seen ____________________?

22. A book contains much ____________________.

23. ____________________ does school close?

24. We gave one book to ____________________ boy.

25. John ____________________ been sick.

26. He gave me ____________________ of them.

27. Is he the man ____________________ sells food?

28. I ______________________________ someone to read it to me.

29. The book ______________________________ I am reading is very interesting.

30. I ______________________________ him coming.

31. I wrote him a ______________________________ letter.

32. ______________________________ you come all the way by foot?

33. There are times ______________________________ one must be early.

34. ______________________________ did it happen?

35. Six and two do ______________________________ make ten.

36. The food ______________________________ eaten.

37. ______________________________ is your name?

38. John thought so, but I ______________________________ agree.

39. He had ______________________________ an important man.

40. ______________________________ man I saw had no feet.

TEST CONSTRUCTION

The construction of the Language Sampling Device (LSD) and the Language Sampling Test (LST) was constrained physically in two ways. First, the investigators were limited by the schools in Japan to two fifty-minute periods per class in which to administer the tests. The investigators tentatively calculated that within a fifty-minute period there would be ample time for instructions and fifty questions. More questions would raise the probability that significant numbers of students might be pressed for time, which would affect their performance negatively. An experimental edition of the LSD of fifty items was, therefore, tried out on three native speakers of English, ages 8, 9, and 9 1/2. They were all able to complete the fifty items in a period of from thirteen to twenty-six minutes. Experimental copies of the fifty-item LSD were also sent to Japan and were tested on three Japanese seventh-graders who were able to complete the test in thirty-five minutes. If Japanese students were able to complete fifty 'cloze' procedure items in thirty-five minutes, it was assumed that other students could easily complete fifty multiple-choice items in the same amount of time or less. This would make the tests power tests in the sense that at least eighty per cent of all students could attempt the fifty items on both tests in two fifty-minute periods.

Secondly, given the fifty-item limit, it would be impossible to obtain data on the total English performance of the Japanese informants. Therefore, the investigators initially decided to limit the tests to measures of Noun-Phrase structures, Verb-Phrase structures, Negative structures, Interrogative structures, and a limited set of lexical items (Bendix, 1966).

Rather than distribute the questions evenly among the five areas of interest, the investigators decided that the Noun-Phrase and Verb-Phrase structures should be more heavily represented in that the various possibilities in these are far richer than in the others. The final assignment is shown in Table C.1.

The specific procedures followed in constructing the LSD and LST included the following steps:

> Step I. Defining Verb-Phrase structures, Noun-Phrase structures, Negative structures, Interrogative structures, and Lexical items in terms of word class categories.

TABLE C.1

Initial Blueprint of the Tests

Verb-Phrase Structures:	15 items
Noun-Phrase Structures:	15 items
Negative Structures:	8 items
Interrogative Structures:	7 items
Lexical Items:	5 items

Step II. Selecting sentences to be used in LSD.

Step III. Writing the LSD items and instructions.

Step IV. Administering an experimental form of the LSD to native speakers of English and editing the items on the basis of their performance.

Step V. Writing the items and distractors for the LST.

Step VI. Reproducing the LSD and LST for wider use in collecting samples from informants.

The details of each of these steps are discussed in turn below.

Each of the structures to be tested was defined in terms of word-class categories. Noun-Phrase structures were defined as Head Nouns, Left-Branching Modification Structures, Right-Branching Modification Structures, and Noun Substitutes. Head Nouns were further defined as Count Noun, singular, Count Noun, plural, and Mass Noun. Left-Branching Modification Structures included Adjectives, Determiners, and Pre-Determiners. Right-Branching Modification Structures were redefined as Relative Pronouns and Noun Substitutes as Pronouns and Function Nouns. Verb-Phrase structures were defined as Main Verbs, Auxiliary structures, and Verb Complement structures. Main Verbs were further subdivided according to formal characteristics: Verb + Ø, + -s, + -ed, + -en, and + -ing. Auxiliary structures were categorized according to the form of the Main Verb they occur before, i.e., Auxiliary + V -Ø, + V -ing, + V -en, or + to V. Verb Complements were defined in terms of the form of the Complement following the Main Verb with or without an intervening Noun Phrase, i.e., Main Verb (±Noun Phrase) + V -Ø, + to V, or + V -ing. Negative structures were defined as any single word having negative meaning, e.g., NOT, NO, NOTHING, NOR, DIDN'T, etc. Interrogative structures were defined as questions containing an interrogative word such as WHAT, WHEN, HOW, etc., or questions without interrogative words but formed by transformation of the Auxiliary before the Subject Noun Phrase. The lexical items were defined according to Bendix (1966) and included: FIND, GIVE, GET, KEEP, TAKE.

For each of the categories defined, an alphabetical list of tokens for each of the categories was obtained from a list of the first 500 words on the Kucera-Francis frequency list (1967), i.e., words ranging from 69,971 to 194 occurrences in a million words. The list of tokens was limited to the first 500 in order to reduce any interference in the testing procedure from vocabulary unknown to the informant, thus making the tests more purely measures of syntactic structures.

The blueprint for the tests was then redefined in number of tokens for each of the word-class categories as shown in Table C.2.

TABLE C.2

Initial Blueprint of the Tests Defined in Word-Class Categories

1. Noun-Phrase Structures: 15 items
    a. Noun - 3 tokens
    b. Adjectives - 3 tokens
    c. Determiners - 3 tokens
    d. Relatives - 3 tokens
    e. Noun Substitutes - 3 tokens
2. Verb-Phrase Structures: 15 items
    a. Main Verbs - 6 tokens
    b. Auxiliaries - 6 tokens
    c. Complements - 3 tokens
3. Negative Structures: 8 items
4. Interrogative Structures 7 items
    a. Wh- Questions - 3 tokens
    b. Regular Questions - 4 tokens
5. Lexical Items: 5 items

Twice the number of required tokens for each category were then randomly selected from the alphabetical lists, using a random numbers table for the numbers 1 - 99. Each of the selected tokens was then defined semantically in terms of its most frequent meaning. Semantic frequencies of the selected tokens were obtained from Michael West's GENERAL SERVICE LIST. A semantic index for each token was obtained by multiplying the number of occurrences on the Kucera-Francis list by the semantic frequency. If the resulting product was greater than 194, the word was defined in that meaning. If not, the word was rejected and another word selected randomly until the number of desired tokens was obtained. The purpose of the semantic index was to further control interference in the tests from unknown vocabulary.

Sentences containing the selected tokens were obtained from the entry for the token in the THORNDIKE CENTURY JUNIOR DICTIONARY. If no sentences could be found, West's GENERAL SERVICE LIST and THE RANDOM HOUSE DICTIONARY OF THE ENGLISH LANGUAGE were each consulted in turn until sentences were obtained for each of the 100 selected tokens.

The Thorndike dictionary was consulted first because nearly every entry has a simple sentence illustrating the use of the word which is comprehensible to most children. The West list was consulted second because its definition of semantic meanings is in illustrative sentence fragments which are simple to understand. The RANDOM HOUSE DICTIONARY was consulted last because its illustrative sentences tend to be more complex.

All of the words in each of the 100 sentences thus obtained were checked against the Kucera-Francis list to see if their frequency characteristics were within the first 500 word range. If not, the word was replaced by a suitable word with the desired frequency characteristics.

The key word of each sentence, i.e., the word through which the sentence was obtained from one of the dictionaries, was deleted. The remaining framework became the stem for a 'cloze' item on one of the two experimental forms of the LSD, each containing fifty items.

The experimental forms of the LSD were administered to sixty-five University of Hawaii freshmen in July, 1970. The items on the experimental forms were scored in the following way:

a) the blank was filled with the key word;
b) the blank was filled with a substitute which matched the key word in terms of word-class category;
c) the blank was filled with a word that was grammatically acceptable, but of a different word-class category;
d) the blank was filled with a word that was not grammatically acceptable.

Any item in which (c) or (d) had more than three entries, i.e., 4.6% of sixty-five was rejected. An error rate of 5% would invalidate the test and make any interpretation of the results obtained by the LSD meaningless in terms of a proficiency test of English syntax.

The number of remaining acceptable items for each word-class category is shown in Table C.3.

TABLE C.3

Number of Items on the LSD Accepted on the Basis of Native Speaker Performance

Structures	Number Accepted	Number Desired
1. Noun-Phrase Structures	27	15
a. Noun	4	3
1. Count Noun, singular	1	1
2. Count Noun, plural	2	1
3. Mass Noun	1	1
b. Adjectives	4	3
c. Determiners	5	3
1. Determiners	4	2
2. Pre-determiners	1	1
d. Relative Pronouns	6	3
e. Noun Substitutes	4	3
1. Pronouns	3	2
2. Function Nouns	1	1
2. Verb-Phrase Structures	14	15
a. Main Verbs	6	6
1. V -Ø	1	1
2. V -s	--	1
3. V -ed	3	2
4. V -en	1	1
5. V ing	1	1
b. Auxiliaries	6	6
1. + V-Ø	1	2
2. + V-ing	1	1
3. + V-en	3	2
4. + to V	1	1
c. Complements	2	3
1. + V-Ø	--	1
2. + to V	1	1
3. + V-ing	1	1

TABLE C.3 (continued)

Structures	Number Accepted	Number Desired
3. Negative Structures	4	8
4. Interrogative Structures	11	7
a. Wh- Questions	3	3
b. Regular Questions	8	4
5. Lexical Items	--	5

As can be seen, the Complements category yielded two acceptable items instead of the desired three, the Negative category yielded four acceptable items instead of the desired eight, and none of the five lexical items proved acceptable. Because of the lack of testable items, the initial blueprint of the test was revised and reduced to forty items. The resulting blueprint is shown in Table 3.1 in Chapter 3 above. For the categories Count Noun, plural, Adjectives, Determiners, Relative Pronouns, Pronouns, and Regular Questions where the number of acceptable items exceeded the number desired, items were chosen for exclusion in the final form of the LSD on the basis of the least number responses by native speakers which filled the blanks with words which were grammatically acceptable, but of a different word-class category or with words which were grammatically unacceptable. The forty acceptable items were then randomized and the instructions were translated into Japanese.

The finished LSD was used as the format for the LST. The same word-class categories and order of items was observed. If for any category, there was a usable item which did not appear on the LSD, it was used as the basis for the LST items. This process yielded twelve of the LST items. For the remaining eighteen items, the LSD item was used and words were substituted in the LSD sentence frame in keeping with the key word which was preserved. In this way, the stems for all of the forty LST items were obtained.

The three distractors for the multiple-choice items on the LST were selected on the basis of the word-class category being tested. If, for example, the category was one of verbal tense, the distractors were various tenses. If, however, this process did not yield a sufficient number of distractors, others were selected either from the errors

made by the University of Hawaii freshmen on the experimental edition of the LSD or on an arbitrary basis. No attempt was made to select distractors on the basis of the fact that Japanese students were going to take the test.

APPENDIX D

OPTIONAL/OBLIGATORY DECISIONS ON THE LSD SENTENCES

Three investigators were given the task of determining whether the filler for each of the LSD sentence items was constructionally optional or obligatory: marked (-) (optional) if the different acceptable fillers would involve different transformational histories, and (+) (obligatory) if the different acceptable fillers would all involve the same transformational history. Ordered marking triplets, (+++), (++-), etc. were thus obtained. It was immediately noticed that the distribution of these triplets might be random.

The probability that each investigator would mark an item (+) was estimated by dividing the number of his (+) markings by 40, with the resulting quotients: $p_1 = .775$, $p_2 = .45$, $p_3 = .62$. (The probability of a (-) marking is, of course $(1-p_i) = q_i$, or .225, .55, and .38 respectively.) The probability of any particular triplet, such as (+-+), was calculated by multiplying the appropriate p_i's and q_i's - in this case, p_1, q_2, and p_3: (.775) (.55) (.62) = .264 - and the expected number of items out of the forty to be marked in this way was obtained by multiplying the obtained probability by 40 - in this case, (40) (.264) = 10.6, the nearest round number to which is 11. In table D.1, the number of items that are predicted by the above calculation to have any particular triplet are listed under _e_ (for "expected"), and the number of actual occurrences of the same triplet is listed under _n_ (for "number").

TABLE D.1

Expected and Actual Numbers of Marking Triplets.

Triplet	e	n	$(e-n)^2/e$
(+++)	9	7	.44
(+-+)	10	11	.10
(++-)	5	8	1.80
(-++)	3	2	.33
(+--)	6	5	.17
(--+)	3	5	1.33
(-+-)	2	1	.50
(---)	2	1	.50
Total:	40	40	5.17

For 6 degrees of freedom (df = r - 1 - g, where r is the number of values, here 8, and g is the number of estimated values that were required in the statistic, in this case 1, estimating p_i), the nearest chi-square value above 5.17 is 5.38, which is significant at the .50 level, indicating that the null hypothesis (that _e_ cannot predict _n_) must be rejected. (Crow, pp. 85-87)[1]

No matter how one of the three investigators may have been motivated to mark obligatoriness, the other two were motivated in ways that are only randomly related to the first and to each other. For this reason, it was decided that the obligatory and optional choices could not sensibly be used in establishing a hierarchy based on the Stockwell model.

---

[1] Edwin L. Crow, Francis A. Davis, and Margaret W. Maxfield, STATISTICS MANUAL, New York: Dover Publications, 1960, pp. 85-87.

APPENDIX E

THE ENGLISH SENTENCES AND THEIR JAPANESE TRANSLATIONS

(Capitalization indicates the deleted word on the LSD and its Japanese equivalent(s))

1. I do NOT believe it.
   Watakushi wa sore o shinjiNAI.

2. I WAS late to school last week.
   Senshuu gakkoo ni okureTA.

3. Have you any money? We haven't ANY.
   Okane ga aru ka? Watakushi wa // nai / motte inai / zenzen ja nai//.

4. Her HOUSE is at 25 South Street.
   Kanojo no UCHI wa 25 South Street // desu / ni aru //.

5. He CAN'T speak English yet.
   Kare wa mada eigo // o hanasENAI / wa DEKINAI //.

6. They ARE waiting.
   Kare wa matte IRU.

7. Give ME a drink.
   Nomimono o (ippai) (WATAKUSHI NI) kudasai.

8. He is READING a book.
   Kare wa hon o YONDE iru.

9. She GAVE a party last night.
   Kanojo wa sakuban paati o // HIRAITA / SHITA //.

10. John is an INTERESTING person.
    John wa OMOSHIROI hito desu.

11. He was to come at 10:00.
    Kare wa 10:00 ni kuru // koto ni natte ITA / hazu DATTA //.

12. WILL you help us?
    Watakushitachi o // tetsudatte KUDASAI / tasukete KUDASAIMASE //.

13. Let's PAY a visit to New York.
    (a) New York e itte miyoo.
    (b) New York ni kembutsu ni ikoo.
    (c) New York o otozuremashoo.

14. DOES she have any?
Kanojo wa // ikuraka / nanika // motte iru KA?

15. You HAVE read this book.
Anata wa kono hanashi o yonde // IRU / SHIMATTA //.

16. I worked there three full YEARS.
Watakushi wa soko de maru sanNEN hataraita.

17. CAN I go now?
Ima itte(mo) II DESU KA?

18. School BEGAN at 9 last year.
Kyonnen wa gakkoo wa 9 ni HAJIMATTA.

19. A child WAS loved by its mother.
Kodomo wa hahaoya (a) kara kawaigaRARETA mono da.
(b) ni aisARETE ITA.

20. I've written a NEW story about it.
Watakushi wa sore ni tsuite ATARASHII hanashi o kaita.

21. She isn't here. Have you seen HER?
(Kanojo wa) inai wa. (a) ANOKO minakatta?
(b) Dokoka de [KANOJO O] mita?
(c) [KANOJO WA] dokoka inakatta ka?

22. A book contains much INFORMATION.
(a) Hon ni wa takusan JOOHOO ga haitte iru.
(b) Hon ni wa iroiro na KOTO ga dete iru.

23. WHEN does school close?
Gakkoo wa ITSU // owaru no deshoo ka / shimaru ka //?

24. We gave one book to EACH boy.
Otoko no ko ni (hon o) issatsu ZUTSU (hon o) ageta.

25. John HAS been sick.
John wa (a) zutto byooki DA.
(b) fusette ORAREMASU.
(c) chooshi ga warukatTA NO DA.

26. He gave me ONE of them.
Kare wa (sono naka no) HITOTSU o itadaita.

27. Is he the man WHO sells food?
Kare ga tabemono o uru hito desu ka?

28. I ASKED someone to read it to me.
(Sore o) (aru hito ni) yonde kureru yoo ni (a) TANONDA.
(b) ONEGAISHITE MITA.

29. The book WHICH I am reading is interesting.
Watakushi ga yonde iru hon wa omoshiroi.

30. I SAW him coming.
(a) Kare ga kuru no o MITA.
(b) Kare ga kuru yo. MITA.

31. I wrote him a LONG letter.
NAGAI tegami o kaita.

32. DID you come all the way by foot?
(a) Norimono o goryoo nasarazu ni irassiTA no desu KA?
(b) Norimono ni naranaide konna tooku made kiTA no KAI?

33. There are times WHEN one must be early.
Hayaku shinakereba naranai toki ga aru.

34. HOW did it happen?
(a) DOOSHITE sonna koto ni nattan' dai?
(b) DOO shitan' dai?

35. Six and two do NOT make ten.
(a) 6 tasu 2 wa 10 ja NAI desu.
(b) 6 ni 2 jaa // 10 nya naraN yo / 10 ni naru wake wa NAI //.

36. The food WAS eaten.
Sono tabemono o tabete SHIMATTA.

37. WHAT is your name?
Anatano namae wa NAN desu ka?

38. John thought so, but I DON'T agree.
(a) John wa soo omotta ga watakushi wa // soo omowaNAI / hantai da yo //.
(b) John no iu koto ni sansei dekiKANEMASU.

39. He had MET an important man.
(a) Kare wa erai hito ni ATTA.
(b) Aru suji no yoojin to KAIKEN shita.

40. THAT man had no feet.
ANO hito wa ashi ga nai.

APPENDIX F

LINGUISTIC ANALYSIS OF THE SENTENCES FOR CONTRAST

Since contrasts were performed on the basis of specific descriptions of the sentences involved, preliminary descriptions of all sentences according to the various different analyses were prepared, except for the Stockwell, for which no formal definitions were available for the variables STRUCTURAL SIMILARITY and FUNCTIONAL/SEMANTIC SIMILARITY.

The Kleinjans, Jackson, and Hashimoto descriptions differed in that in Kleinjans and Jackson, the FORM, MEANING, and DISTRIBUTION of the filler only were described, and in the Hashimoto a transformational history of the whole sentence was indicated, as well as an estimate of the translatability of the two sentences.

An example of each analysis is given below, on the first English and Japanese sentences.

English: I do not believe it (NOT deleted on the LSD)
Japanese: Watakushi wa sore o shinjinai (NAI "equivalent" to NOT)

Kleinjans

English	Japanese
FORM: NOT; (-) bound, AUX__V, (-) inflected	FORM: NAI; (+) bound, V__, (+) inflected
DISTRIBUTION: Contrasted forms, shifts with auxiliary	DISTRIBUTION: Can be predicate
MEANING: Negation	MEANING: Negation, tense

Jackson

English	Japanese
FORM:	FORM:
a. major sentence (U)	a. major sentence (U)
b. predicate level (P)	b. dooshibu level (D)
c. auxiliary sector (X)	c. verbal sector (V)
d. negator sub-sector (Neg)	d. no sub-sector
e. filler = negator (listed lexeme class): NOT (neg)	e. filler = adjectivex (a-x) of the dooshi (nonlisted lexeme class), shinji + negative adjective (listed lexeme class) NA-i.

APPENDIX F (cont.)

DISTRIBUTION: Shiftable to preposed auxiliary sector ($\overline{X}$) with the carrier substitute before Subject sector (S) with change to question.	DISTRIBUTION: Nonshiftable.
MEANING: Modification (unilateral dependence on auxiliary; negative status.	MEANING: Predication (mutual dependence on shugobu); negative, non-past indicative status.

Hashimoto

NP AUX Vt NP -- t --→ NP ga NP o Vt

Tneg: NPtenseVP ---→ NPtense+notVP	Tneg: XVPZ ---→ XVP+nak-1Y

TABLE OF X-VALUES (PREDICTIONS OF LEVEL OF DIFFICULTY)

The predictions of level of difficulty for each item on the LSD were given numerical values as follows:

Kleinjans: Level A: 1, Level B: 2, Level C: 3, Level D: 4.
Jackson: Level A: 1, Level B1: 2, Level B2: 3, Level B3: 4, Level C1: 5, Level C2: 6, Level C3: 7, Level D: 8.
Hashimoto: The same values as the numerical value of the hierarchy as determined by each investigator (see section 5.2).
Stockwell: The same values as the numerical value of the hierarchy.

In Table G.1, the column headed Item No. refers to the number of the LSD item. Under each of the analyses, the columns headed 1, 2, 3, 4, represent the predictions made by the different investigators, and the column 5(H) represents the predictions attributed to Hashimoto (see section 5.2). The columns headed a represent the "average" predictions, or the prediction obtained when two or more of the investigators agreed on a level of difficulty (note that there is no a column for the Hashimoto analysis).

TABLE G.1

Predictions of Level of Difficulty

Item No.	Kleinjans				Jackson				Hashimoto					Stockwell			
	1	2	3	a	1	2	3	a	1	2	3	4	5(H)	1	2	3	a
1	3	3	3	3	8	8	8	8	3	4	7	3	1	4	4	1	4
2	3	3	2	3	7	7	7	7	3	2	4	4	1	4	5	4	4
3	4	4	4	4	6	4	1	-	4	2	7	2	1	5	2	5	5
4	1	2	2	2	1	4	1	1	0	0	0	1	-	1	1	-	1
5	3	3	2	3	8	8	8	8	3	4	7	4	3	5	4	1	-
6	1	2	1	1	7	7	5	7	1	1	3	2	0	1	1	2	1
7	2	3	3	3	6	6	1	6	3	3	7	3	-	4	1	-	-
8	1	2	2	2	1	4	1	1	1	1	3	2	0	1	1	1	1
9	1	2	1	1	1	1	1	1	1	1	3	2	1	1	1	2	1
10	1	2	2	2	6	-	1	-	0	0	1	1	0	1	4	1	1
11	4	4	2	4	8	8	8	8	3	1	7	4	2	3	4	-	-
12	2	3	2	2	8	8	8	8	3	4	7	3	3	4	5	-	-
13	3	4	3	3	1	4	6	-	5	3	7	5	-	5	4	-	-
14	3	3	2	3	7	8	7	7	3	3	5	4	3	5	5	1	5
15	3	4	2	-	7	7	5	7	3	3	7	4	2	5	4	-	-
16	2	3	3	3	4	4	6	4	2	1	7	3	-	1	4	1	1
17	4	4	3	4	8	8	8	8	3	4	7	4	3	4	4	2	4

TABLE G.1 (continued)

Item No.	Kleinjans				Jackson				Hashimoto					Stockwell			
	1	2	3	a	1	2	3	a	1	2	3	4	5(H)	1	2	3	a
18	1	2	1	1	6	4	1	-	1	1	3	2	1	1	1	1	1
19	3	4	3	3	8	8	8	8	1	3	6	2	1	4	4	4	4
20	1	2	1	1	6	4	1	-	0	0	1	1	0	1	4	1	1
21	2	3	2	2	7	8	5	-	5	3	4	6	-	4	1	-	-
22	3	4	3	3	4	4	4	4	3	3	6	4	-	4	4	4	4
23	1	2	2	2	3	8	7	-	2	0	1	2	0	1	1	2	1
24	4	3	3	3	8	4	7	-	3	3	5	4	-	2	3	1	-
25	3	4	3	3	-	-	-	-	3	3	8	4	1	5	5	-	5
26	2	3	1	-	6	4	1	-	5	3	6	4	3	1	1	-	1
27	2	3	3	3	4	4	6	4	5	3	4	4	2	5	5	-	5
28	2	2	1	2	1	-	1	1	3	4	4	4	2	4	1	4	4
29	2	3	3	3	2	4	6	-	5	3	6	4	2	5	5	1	5
30	2	2	1	2	-	-	1	-	3	4	5	4	3	4	1	4	4
31	1	2	1	1	6	4	1	-	0	0	1	1	0	1	4	4	4
32	2	3	2	2	8	8	8	8	3	3	7	4	3	5	5	4	5
33	2	3	3	3	4	4	6	4	5	3	7	4	2	5	5	4	5
34	1	2	1	1	3	-	5	-	0	0	7	1	0	1	4	4	4
35	3	3	2	3	4	3	8	-	3	3	5	4	1	4	4	4	4
36	4	3	2	-	7	8	5	-	6	4	8	4	1	4	4	4	4
37	2	4	1	-	4	4	4	4	2	3	8	3	0	1	1	4	1
38	2	4	2	2	4	8	8	8	3	3	5	4	0	5	4	4	4
39	3	3	2	3	1	-	6	-	3	3	7	4	1	4	4	4	4
40	1	2	2	2	1	1	1	1	0	1	8	1	0	1	1	4	1

APPENDIX H

TABLE OF e-VALUES (GROSS PREDICTIONS OF DIFFICULTY)

Table H.1 is derived directly from Table G.1 (Appendix G), the Table of predictions of level of difficulty. The lowest level of difficulty predicted by each investigator under each analysis is considered a prediction of "no difficulty", such that e = 0. Otherwise, difficulty is predicted, and e = 1.

TABLE H.1

e-values (predictions of gross difficulty)

Item No.	Kleinjans				Jackson				Hashimoto					Stockwell			
	1	2	3	a	1	2	3	a	1	2	3	4	5(H)	1	2	3	a
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	0	-	1	1	1	1	1	1	1	1	1
4	0	0	1	1	0	1	0	0	0	0	0	0	-	0	0	-	0
5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	-
6	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	1	0
7	1	1	1	1	1	1	0	1	1	1	1	1	-	1	0	-	-
8	0	0	1	1	0	1	0	0	1	1	1	1	0	0	0	0	0
9	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	1	0
10	0	0	1	1	1	-	0	-	0	0	1	0	0	0	1	0	0
11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	-
12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	-
13	1	1	1	1	0	1	1	-	1	1	1	1	-	1	1	-	-
14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
15	1	1	1	-	1	1	1	1	1	1	1	1	1	1	1	-	-
16	1	1	1	1	1	1	1	1	1	1	1	1	-	0	1	0	0
17	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
18	0	0	0	0	1	1	0	-	1	1	1	1	1	0	0	0	0
19	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
20	0	0	0	0	1	1	0	-	0	0	1	0	0	0	1	0	0
21	1	1	1	1	1	1	1	-	1	1	1	1	-	1	0	-	-
22	1	1	1	1	1	1	1	1	1	1	1	1	-	1	1	1	1
23	0	0	1	1	1	1	1	-	1	0	1	1	0	0	0	1	0
24	1	1	1	1	1	1	1	-	1	1	1	1	-	1	1	0	-
25	1	1	1	1	-	-	-	-	1	1	1	1	1	1	1	-	1
26	1	1	0	-	1	1	0	-	1	1	1	1	1	0	0	-	0
27	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	1
28	1	0	0	1	0	-	0	0	1	1	1	1	1	1	0	1	1
29	1	1	1	1	1	1	1	-	1	1	1	1	1	1	1	0	1
30	1	0	0	1	-	-	0	-	1	1	1	1	1	1	0	1	1
31	0	0	0	0	1	1	0	-	0	0	1	0	0	0	1	1	1
32	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1